DAVID LYON '71
122 LOCKHART
452-8987

Stat 191

Plots of Relationship

		Page
Introduction		2-1
2A.	Graphical display	2-2
2B.	Powers and logs	2-6
2C.	Plotting the population of the U.S.A.	2-13
2D.	Plotting the ratio of births to deaths	2-28
2E.	How far have we come?	2-38

******

2K.	How to count	2-40
2L.	How to plot better	2-40
2M.	How to plot without graph paper	2-46

******

2P.	IEDA processor	2-48

******

2R.	How to think about logs	2-49
2S.	How to get low-precision logs easily	2-50
2T.	Critical tables of other handy powers	2-52

******

2X.	Examples	2-56
2Y.	Problems	2-64

Location of tables and plots for 2

Number	Page	Number	Page
1	2-3	21	2-37
2	2-9	22	2-41
3	2-10	23	2-51
4	2-11	24	2-53
5	2-12	25	2-54
6	2-14	26	2-55
7	2-15	27	2-57
8	2-17	28	2-58
9	2-19	29	2-60
10	2-20	30	2-61
11	2-22	31	2-62
12	2-23	32	2-63
13	2-25	33	2-65
14	2-26	34	2-66
15	2-27	35	2-67
16	2-30	36	2-68
17	2-31	37	2-69
18	2-33	38	2-70
19	2-34	39	2-73
20	2-35	40	2-74
		41	2-75
		42	2-76

2

Plots of Relationship

A natural and insightful reaction to much of the discussion of chapter 1 is:

- Why waste time with arithmetic -- graphs would have done all that quicker and easier.

Much can be said for such a view, including:

- if those things attempted in chapter 1 made up all of data analysis, graphs and no arithmetic would be easier.

- when they can be used, graphs do convey more than arithmetic -- more in impact and more in diversity.

But conversely, there are important things to be said in the opposite direction:

- there is no substitute for the combination of graphs and arithmetic.

- there are many problems of data analysis where much arithmetic has to be used to break things out before graphs can show them to us.

- much of what we were carefully explicit about in our arithmetic still occurs, perhaps tacitly or in the mind's eye, with graphs.

In view of all this, we can recognize two important reasons why we took up arithmetic first:

1) we will need generalizations and analogs of the simple arithmetic of chapter 1 as soon as we come to look at data of even slightly more complex structure.

2) taking arithmetic first forces us to concentrate on how it operates.

Here in chapter 2 we shall try to get graphing rolling with equal vigor, so that we can use graphs and arithmetic together, each for what it is best.

We do not wish to stress arithmetic. Arithmetic is a powerful tool, but ought to be thought of as a necessary evil. It is needed to do many things. Because a modern computer can do much arithmetic cheaply and easily, we can be much more tolerant of arithmetic than we once could. Today arithmetic can, if we are fortunate, be only slightly evil.

Arithmetic, wisely used, can do much to break through into the significant areas of our data. If the facts are simple enough, it may even manage to communicate these facts to us. In general, however, we must look to pictures as the most effective way of communicating quantitative information that is at all complex. There is usually no substitute for the friendly graph.

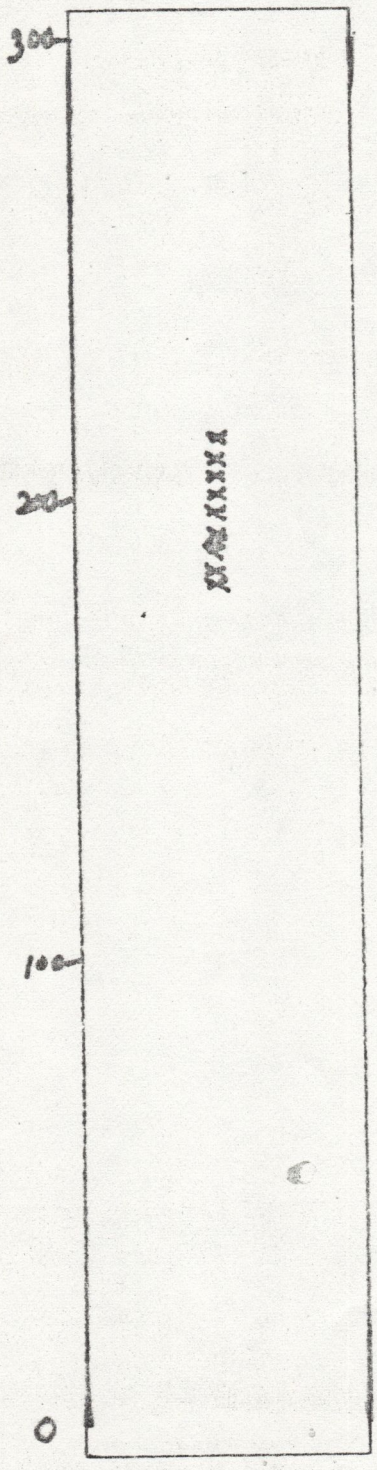
2A. Graphical display

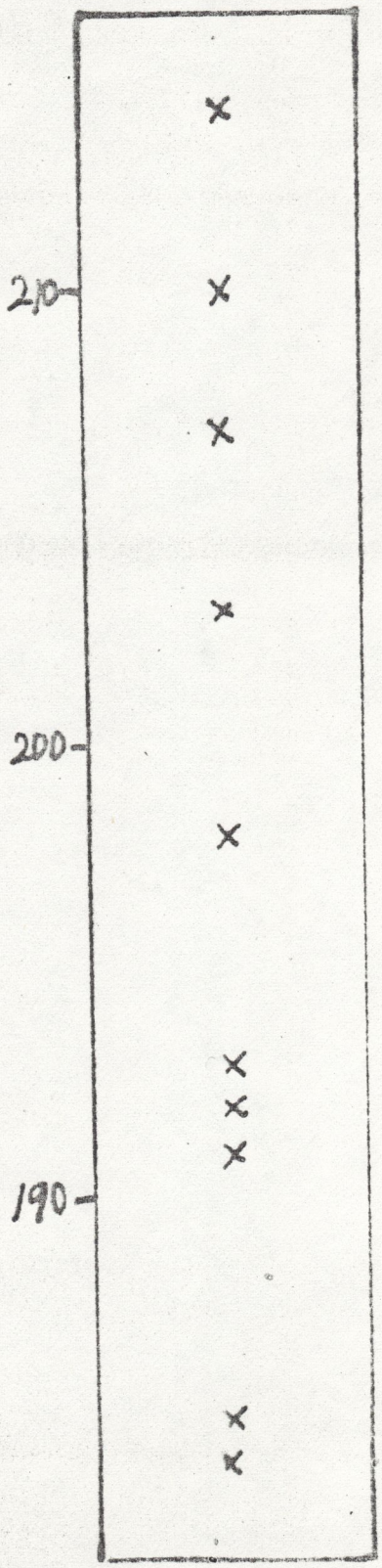
Let us return to the ten bowling scores of chapter 1. Graph 1 shows them in two versions. On the left we have plotted them in a box which starts at zero, on the right in a magnified box that starts at 182. The most obvious fact is that these two pictures serve very different purposes. We need to become quite clear as to what these different purposes are and how they are served by different displays.

The left-hand version shows -- rather dramatically -- that our bowler is getting about 2/3 of a perfect score. It

graph 1 of chapter 2

Two representations of the 10 bowling scores





is saying -- with more impact, but using much more paper -- this bowler is "scoring about 200 out of 300 with most scores reasonably close to 200".

The right-hand version is focused on the spread of the bowler's scores. You can get a very good idea of his typical score -- better than from the left-hand version -- but to do this you must read the scale . When you do, eye and mind will cooperate on an answer somewhere between 195 and 200 -- in the range that mean, median, and midmean had earlier staked out as a matter of arithmetic. The facts about typical value are there for anyone who wants access to them, but what strikes the eye most heavily in the right-hand version is information about the amount and pattern of spread of the scores. We are quite likely, for instance, to notice that higher scores seem more spread out from one another than do lower scores -- to do this even if we had not come to our picture with such a possibility in mind.

One of the great features of graphical impact -- one of the reasons which arithmetic can never serve ALL the functions of graphs -- is the graph's ability to show you that you were not looking for -- that you may even have been very reluctant to see. Examples of this are commonplace; we shall try to show several as we proceed.

In chapter 1 we talked about separating our data in the simplest way -- into typical values and residuals. To do this by arithmetic is routine, detailed, perhaps tedious

(unless an automatic computer can do it for us in seconds). We have really done much the same thing in Graph 1: the left-hand version tells us about typical value, the right-hand version tells us about residuals. Separation is still a major concern, though with graphs it is easier to fail to notice that it is happening. This is particularly so since the right-hand picture tells us almost all about residuals without having chosen a typical value; arithmetic would have to choose such a value, a value that could then be subtracted to make explicit residuals. The residuals in the right-hand picture are implicit, flexibly defined, provided by a mechanism easy to overlook. Yet their message pounds in on us whether or not we like it.

Choosing which graph we plot or look at is very important. Most books or articles on graphical presentation assume that our task resembles showing the obvious to thick-headed viewers. Accordingly they emphasize getting across typical values visually and have much to say about the importance of starting scales from zero. If the graph is needed to convey that 200 is about 2/3 of 300, this is exactly the right view. (For those newspaper readers least familiar with numbers, this may be exactly the challenge.) But if the graph is to meet other needs, it needs to be otherwise constructed.

We are in the detective business. The picture that shows the big desk, the oriental rug on the floor, and the three windows can help the investigator of crime, but the

close-up view of the match on the rug, the scratch on the desk, or the fingerprint on the third window is at least equally important. We will not shun the macrograph in our data analysis, but we will have to make special use of micrographs that enlarge things until we see them clearly.

Drawing a graph need not be difficult, but you can make it hard if you wish. Section 2L explains a few of the most important things that make graphs easier and more effective.

2B. Powers and logs

Numbers do not always come to us ready to plot. Some modification will be needed often, perhaps usually. Changing y to 3y + 12 or x to 2x - 9 is not the kind of a change that matters here. Relabeling the scales of the graph will take care of such changes. In graphing, as in much of data analysis, such changes -- often collectively called "linear coding" -- are inessential.

The simplest important change in how something is expressed in numbers is to replace the initial numbers by a simple power of them. It is not usual for this power to be a square or a cube. Because of how people choose to make instruments and write numbers down, we much more frequently change them in the opposite direction, as by taking:

- square roots ($\frac{1}{2}$ powers).
- reciprocals (-1 powers).
- reciprocals of square roots ($-\frac{1}{2}$ powers).

Hand computation of any of these is much easier using tables -- this would even be true of squares, to say nothing of cubes -- especially since these tables can often be very compact. Sections 2S and 2T will, in due course, explain how such tables can be compact, and how to use such compact tables.

A look at the powers we have already mentioned:

$$-1, \ -\tfrac{1}{2}, \quad +\tfrac{1}{2}, \ +1$$

where the last entry refers to the raw data itself, reveals an obvious gap. What about the zero power?

We were almost all once taught that anything raised to the zero power was 1. Our teachers were not wrong. But that does not mean that there can be nothing to fill the gap. It turns out that the role of a zero power is for these purposes neatly filled by the logarithm -- which we shall familiarly write as "log" as everyone speaks of it.

Some readers may like a review of the few simple things they need to know about logs. This they can find in Section 2R. There are different kinds of logs, but this makes little difference to us here. Of two kinds, each must be a constant multiple of the other, and we have decided that constant multiples can always be taken up in our scales. We will thus write merely "log", except in the very few instances where it makes a difference. (When we need to emphasize that we are using "common logs" -- logs to the base 10 -- we will write "$\log_{10}$".)

Plot 2 shows how $y = (\text{const}) \cdot (\log x)$ fits neatly in among the simple powers. Notice that negative powers need negative signs to make y increase as x increases.

Some may feel that they would learn more by seeing the scales corresponding to these various functions of x, rather as if these functions appeared on a slide rule or in graph paper. Plot 3 meets their wishes.

In plot 3, we see the negative powers on the left, with their closed upper end (for $x = \infty$) moving up more and more rapidly as we approach the logarithm. On the right, the positive powers behave in a mirror image way with a closed lower end (for $x = 0$) moving lower and lower as we approach the logarithm from that side.

If we look across from one scale to another, following the ticks going with a given number - perhaps 1.4 or .7 or more easily 2 or .5 - we see how smoothly and neatly the logarithm fits in between negative powers and positive powers.

The central point is that moving up the scale -- whether from x to x^2 to x^3 or from $-1/x^2$ to $-1/x$ to log x -- corresponds to emphasizing differences among larger x's in comparison with differences among smaller x's. To do the opposite, emphasizing more strongly differences among smaller x, we need to move down the scale - as from x to log x to $-1/x$.

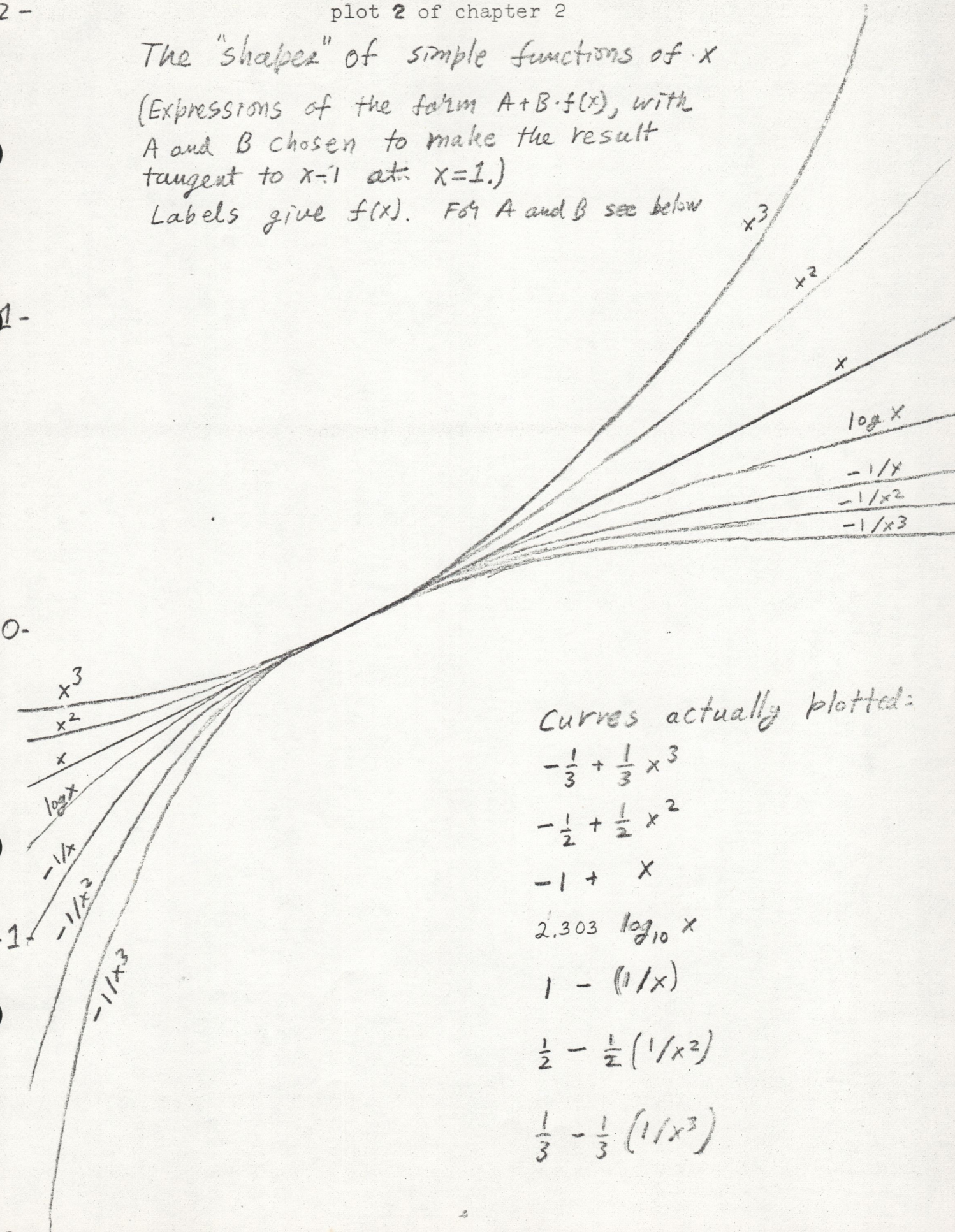
In reexpressing data, we ordinarily work with a narrower range of powers than those just illustrated. Not +3 to -3 but +1 to -1 is what we usually cover. Plots 4 and 5 show similar information for the narrower range. All the same

plot 2 of chapter 2

The "shapes" of simple functions of x

(Expressions of the form $A+B\cdot f(x)$, with A and B chosen to make the result tangent to $x-1$ at $x=1$.)

Labels give $f(x)$. For A and B see below



Curves actually plotted:

$-\frac{1}{3}+\frac{1}{3}x^3$

$-\frac{1}{2}+\frac{1}{2}x^2$

$-1+x$

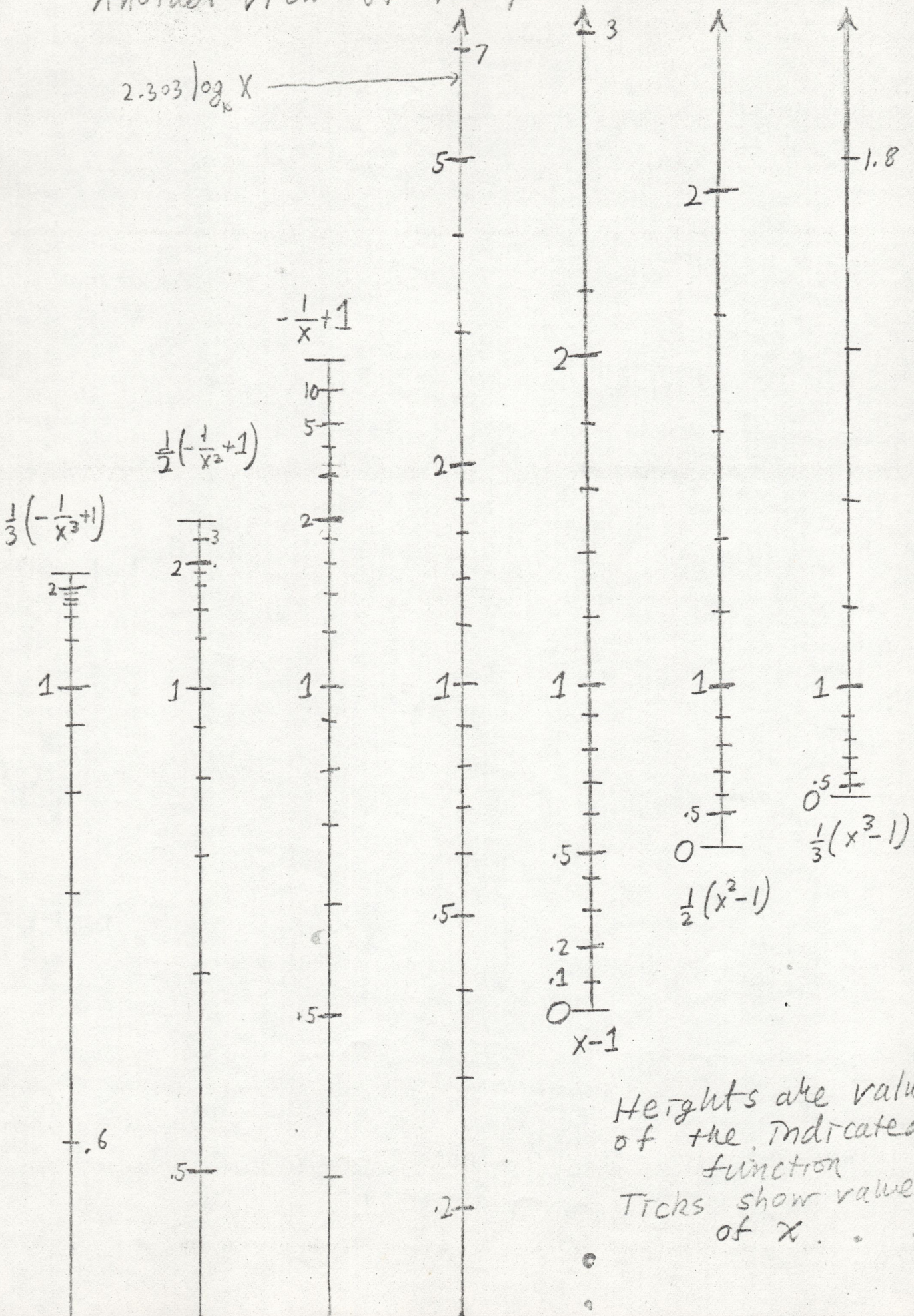
$2.303\ \log_{10} x$

$1-(1/x)$

$\frac{1}{2}-\frac{1}{2}(1/x^2)$

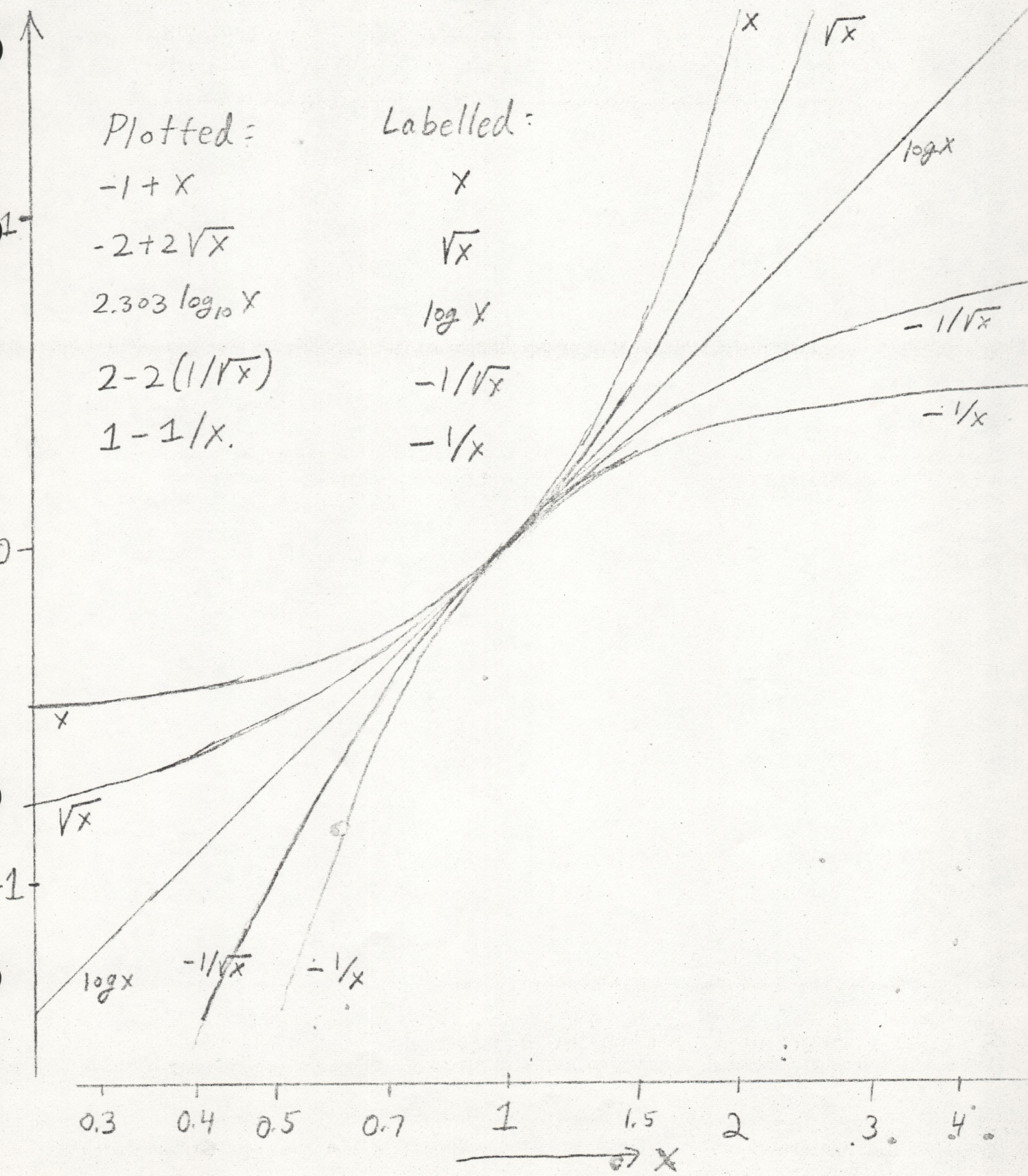
$\frac{1}{3}-\frac{1}{3}(1/x^3)$

Another view of the powers and of logs

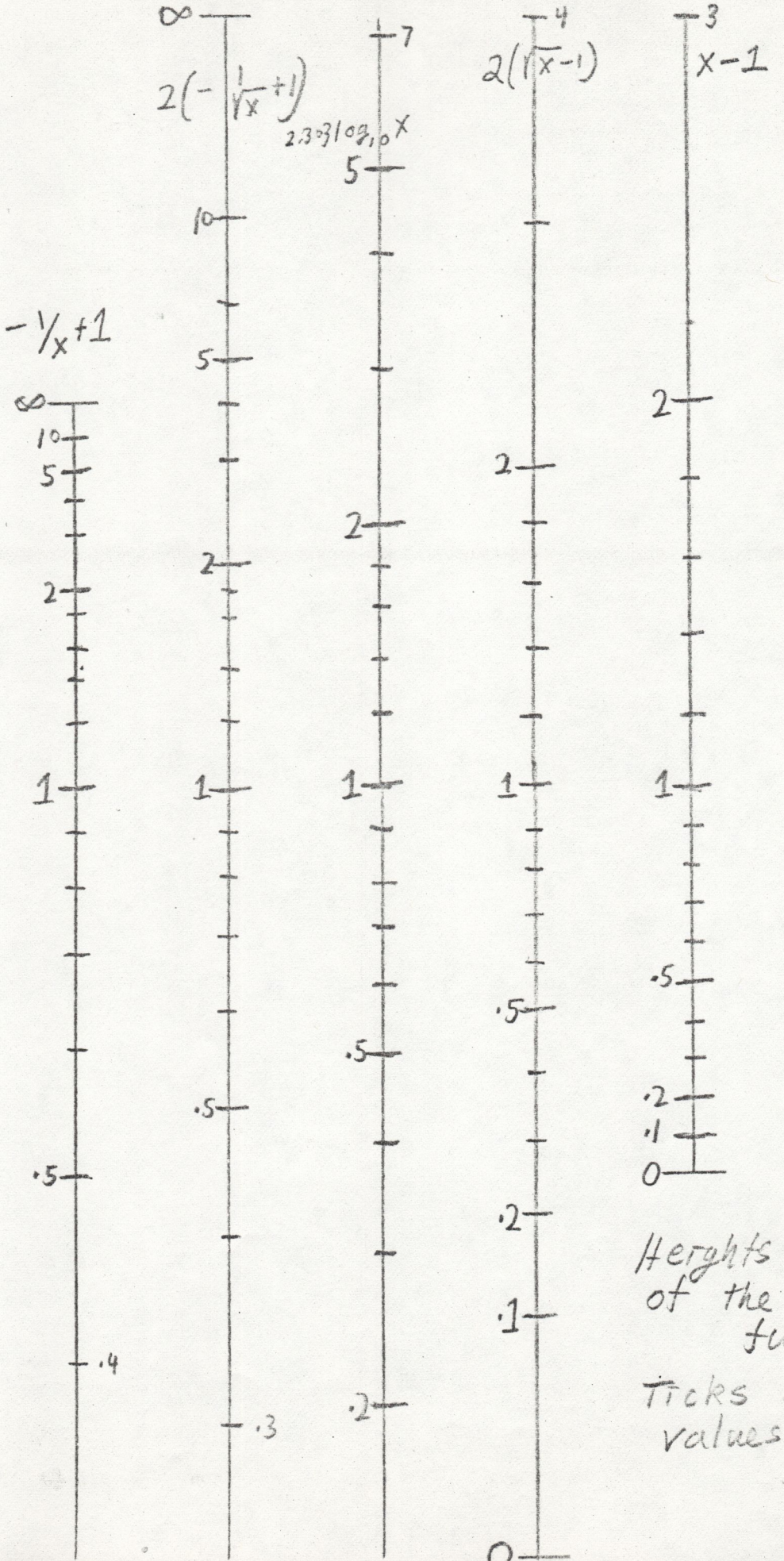


Heights are values of the indicated function
Ticks show values of x.

Logs and the usual powers



The second view of the usual powers and logs



Heights are values of the indicated function

Ticks show even values of x

remarks apply. (Note that the horizontal scale for plot 4 was chosen to make its appearance clear.)

Everything we have been looking at works neatly so long as x is positive, and, even, for positive powers, when x = 0. What about negative x's? We learned long ago not to take square roots of negative numbers. While we could evade this prohibition, we need not. If we choose fractional powers of the form

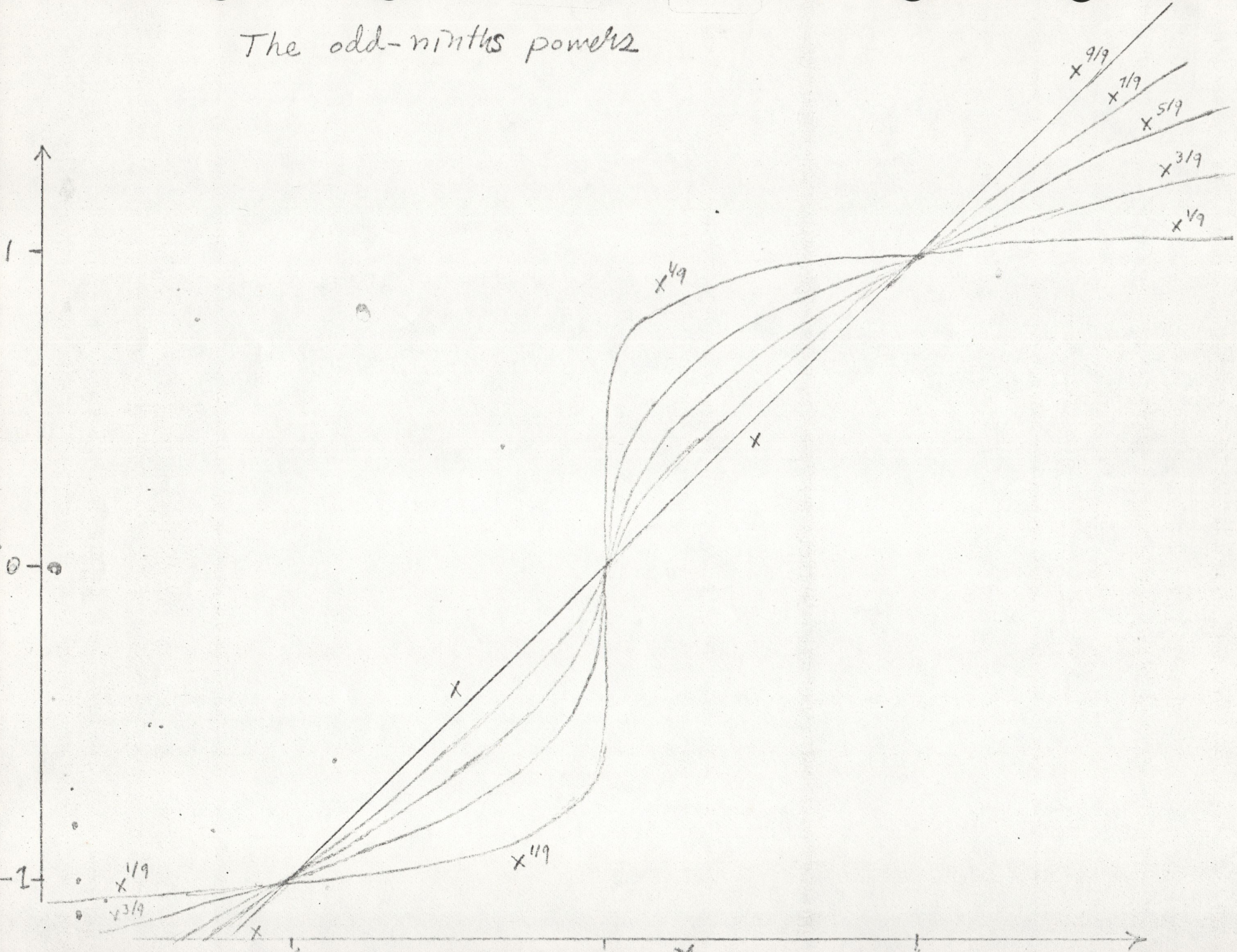
$$\frac{\text{odd integer}}{\text{another odd integer}}$$

we can deal quite comfortably with negative x's, finding negative results. Thus 3/5 or 5/9 powers will play most roles for which square roots might be thought of, and 1/9 or 1/11 powers will play most roles for which logs might have seemed good. Negative powers are rarely in order when x can be negative. Plot 6 shows the behavior of 1/9, 3/9, 5/9, 7/9 and 1st powers across zero. Plot 7 shows how log x behaves when plotted against $x^{1/9}$. The plot cannot be straight, but the curvature is gentle -- to many surprisingly so. And, as we should expect, $x^{1/9}$ comes most of the way from x and $\sqrt{x}$ toward log x.

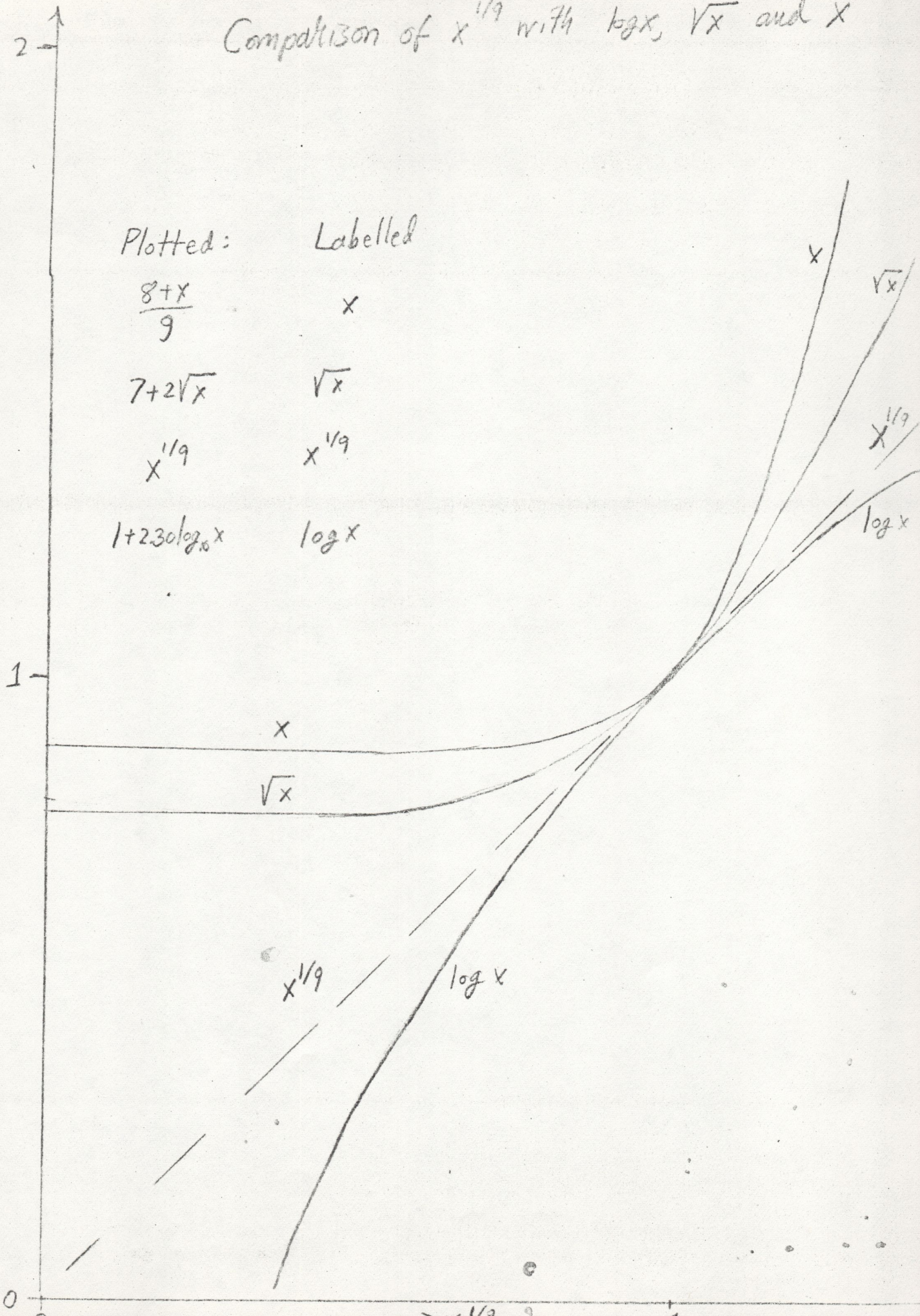
2C. Plotting the population of the U.S.A.

Undoubtedly, the form of graphical representation we have been most exposed to in school is one involving two variables called x and y, in which y is said to depend on x. In data analysis such a plot may help us when we know nothing about the logical connection from x to y -- even whether or not there is one.

The odd-ninths powers



Comparison of $x^{1/9}$ with $\log x$, $\sqrt{x}$ and x



Many people would think that plotting y against x in such a situation is something requiring little thought. If one only wishes to learn little, it is true that a very little thought is enough, but if we want to learn more, we must think more.

A convenient example of this is given by the population of the U.S.A. as counted by the census every ten years from 1800 to 1950. Plot 8 shows the "little thought" version, in which millions of people are plotted against date. What can we see in this graph?

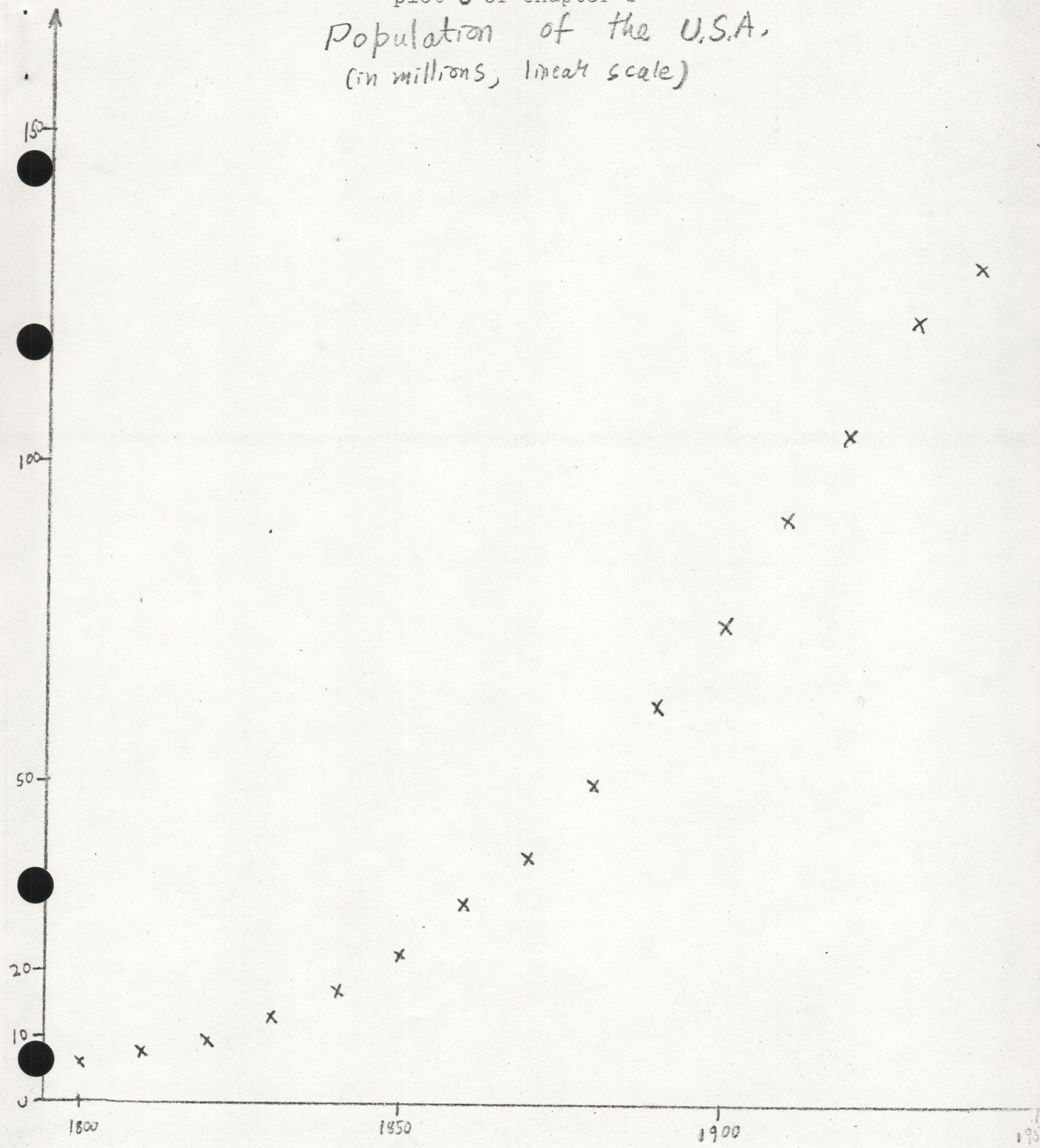
For something like the first half of the period the curve is hollow upward, so that over this period the population was growing at a steadily increasing rate. In fact, it might have been growing at something like a constant percentage each decade. For something like the second half of the period, the growth seems to approximate a straight line. Beyond this, the value for 1940 seems to be somewhat low.

All this is helpful. If we had never before looked at the population of the United States as a function of time, we would have rightly felt that we had learned quite a lot from Plot 8. But once we have come this far, must we stop? Can we not use what we have so far learned to help us look more deeply into the growth of the U.S.A's population? Surely we ought to be able to do this.

In the present instance this is easy to do. What do we have as bases for further steps? Two things come in sight in Plot 8:

plot 8 of chapter 2

Population of the U.S.A.
(in millions, linear scale)



- the early years were years of accelerated growth, possibly at a constant percent per year.

- in the later years, the population grew by about the same number of people each decade.

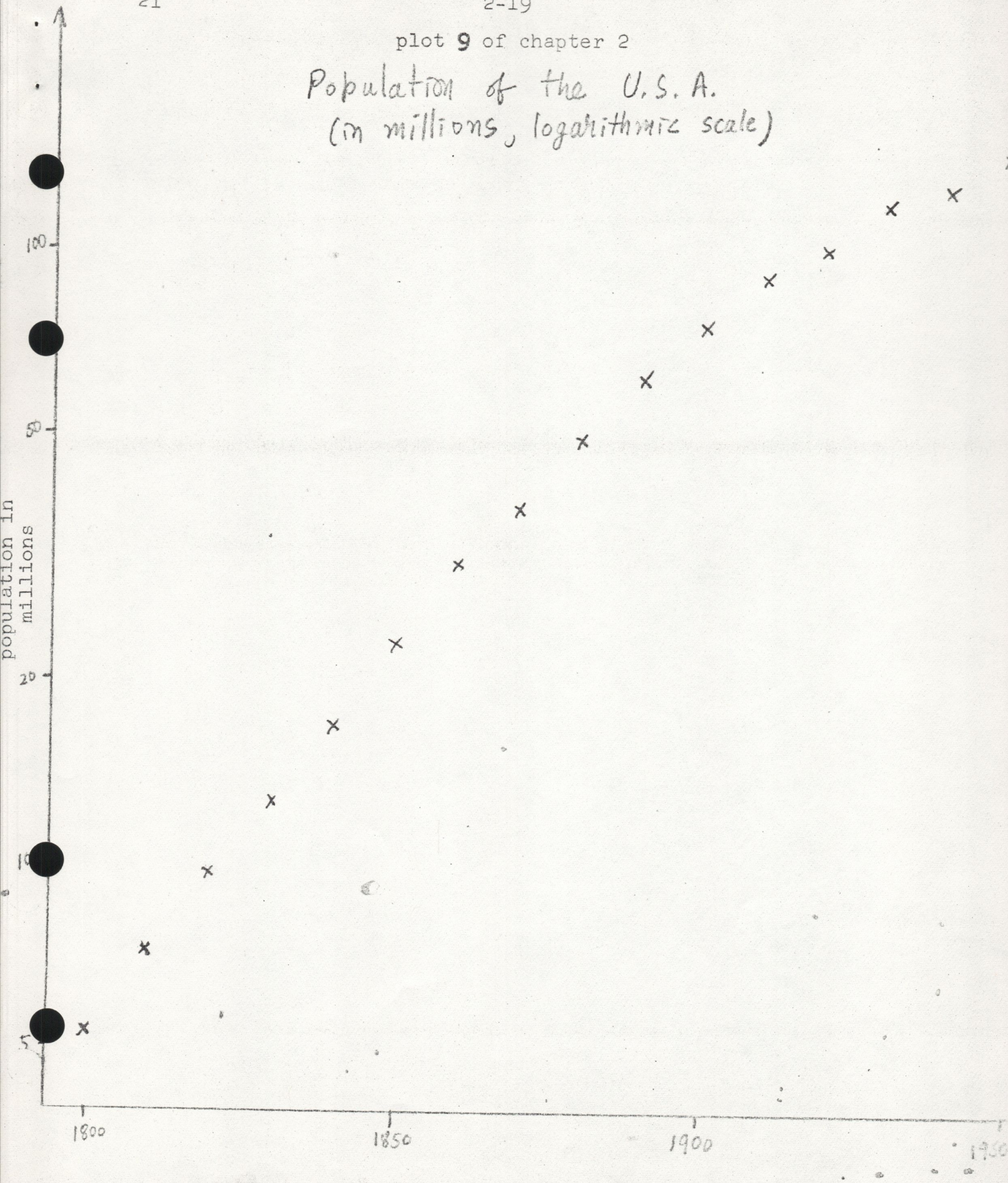
We can check up on these appearances and, more importantly, try to use them to go further.

To check up on constant percent growth, the easy thing is to look up the logarithms of the population sizes, and then plot them against date. (If we didn't expect to go further, we could just plot the raw values on semilogarithmic graph paper instead.) Plot 9 shows the result of doing this. The earlier part of the plot now looks quite straight, even if we put our eye close to the paper so that we can look right along this hypothetical line. This looks very much as if constant percentage growth per decade was a good description of U.S.A. population growth in the early 1800's. Let us record this appearance, and plan to come back to it.

Before going further on that line, however, we shall turn to the apparent linearity of population growth seen in the original graph, plot 8, for the later decades. Plot 10 shows the result of drawing in a comparison line. This line follows the data quite well, leaving us with a confirmed feeling about both approximate straightness and the dip in 1940. Need we stop here? Surely not. The straight line is a partial description of how the data behaves in later years. One of

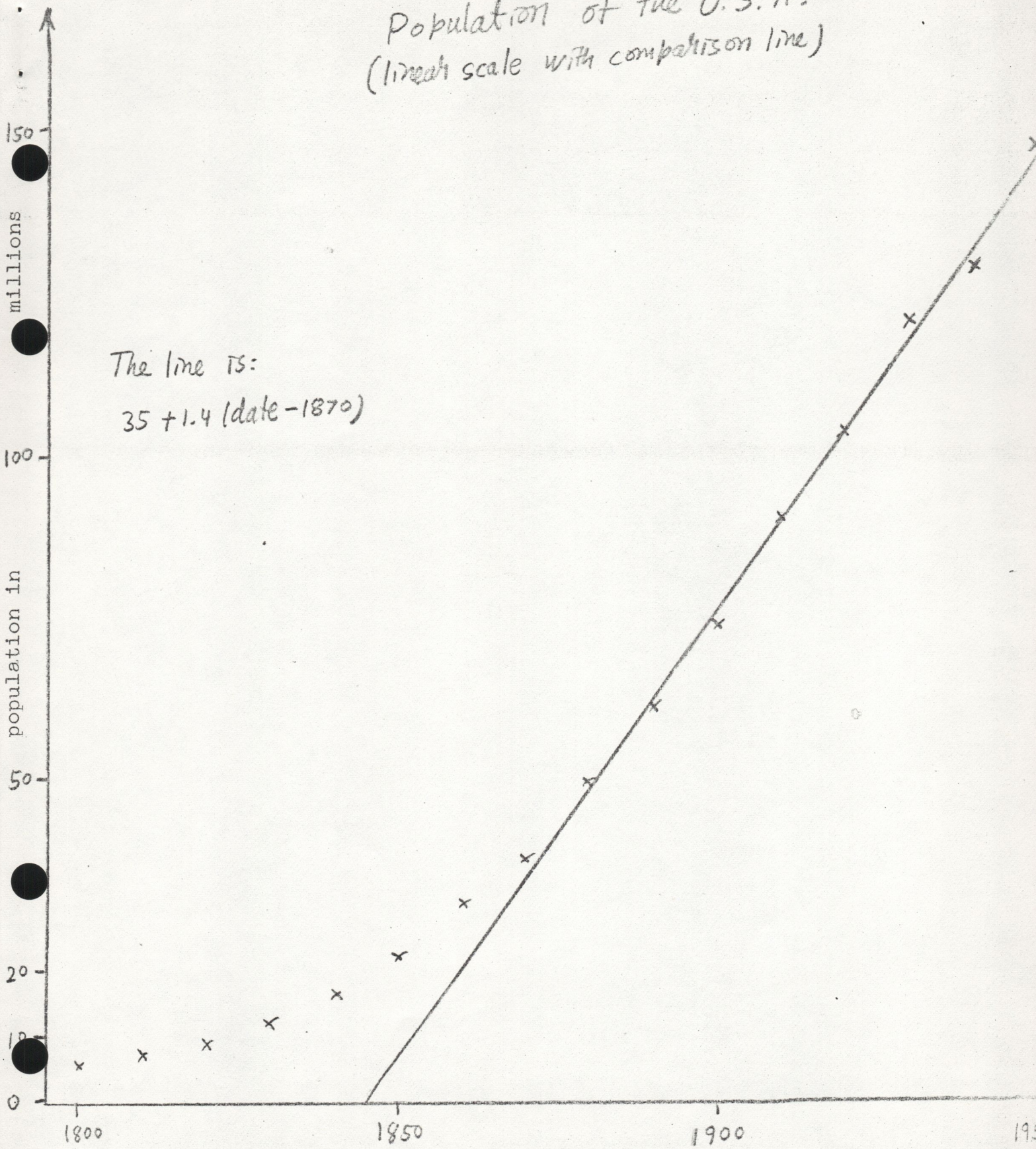
plot 9 of chapter 2

Population of the U.S.A.
(in millions, logarithmic scale)



plot **10** of chapter 2

Population of the U.S.A.
(linear scale with comparison line)



the great arts of data analysis consist of subtracting out of partial descriptions and examining the residuals that are left. Let us do just this.

Plot **11** shows the residuals from the line of plot 10, for the period from 1800 to 1950. The detailed behavior of the population in the later half of this period is now fairly well revealed. The earlier half of the period is, however, telling us little. (Especially since the comparison line

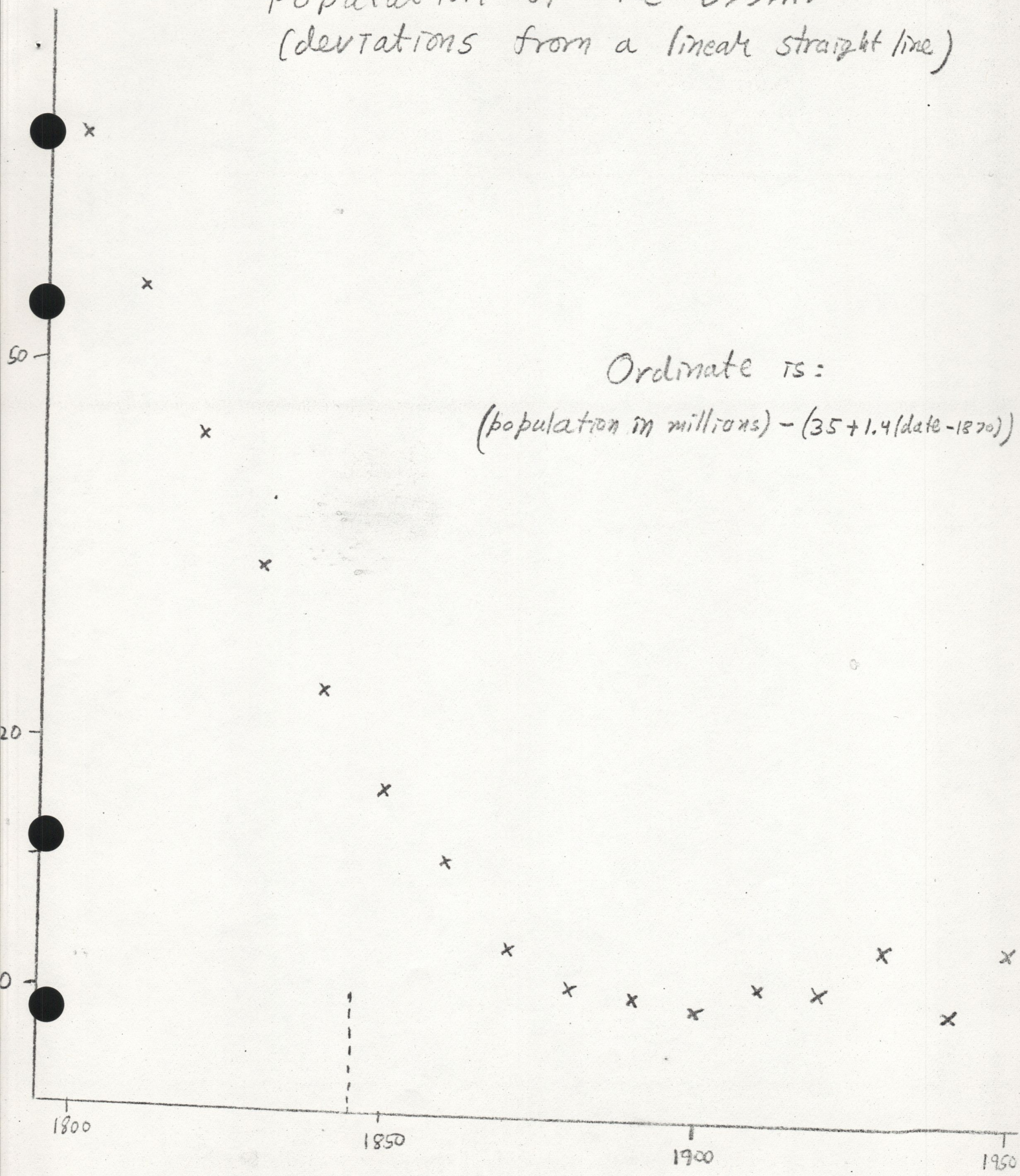
$$35 + 1.4(\text{date} - 1870)$$

gives negative populations before 1845, making the comparison in earlier years unlikely to be helpful.) If we gave up looking at the early half, and concentrated on the later half, we could put the data under a very much more powerful microscope. Why not do just this?

Plot **12** shows the same values as the right-hand side of plot 11, but at 15 times the vertical scale. We can now see that the 1940 population was about five million less than would fit smoothly into the adjacent values. (Why do you think this happened?) And more, for we can now see that 1920 was a couple of millions low also (unless 1930 is thought to be unduly high).

Our magnifying glass is now working at full capacity, at least until we identify some further partial description and arrange to subtract it out also. To learn more about the later years of U.S.A. population growth, we would need either to get year-by-year estimates or to study the mechanisms which are involved.

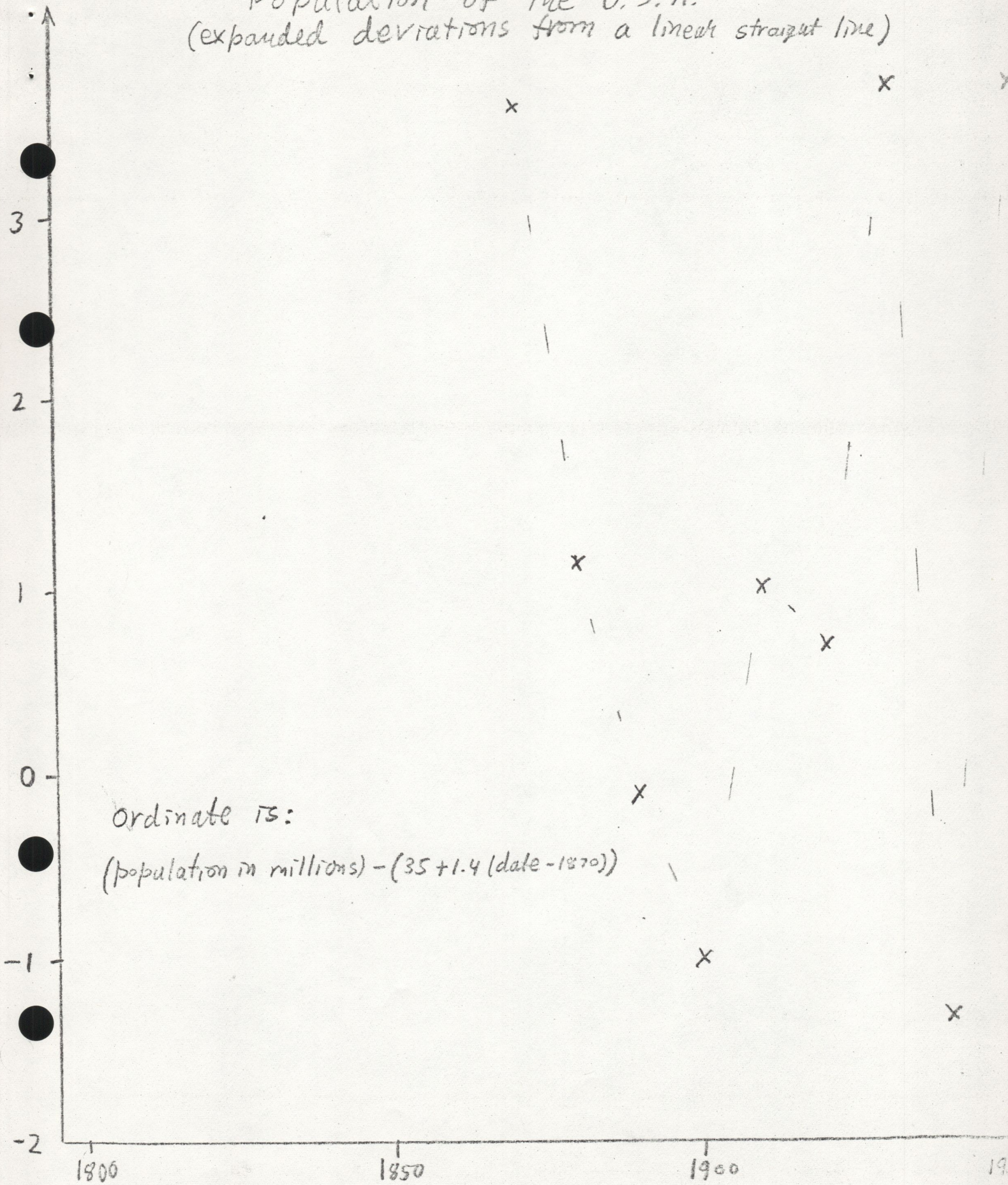
Population of the U.S.A.
(deviations from a linear straight line)



Ordinate is:
(population in millions) − (35 + 1.4(date − 1870))

plot 12 of chapter 2

Population of the U.S.A.
(expanded deviations from a linear straight line)



So much for the later years, what of the earlier ones? We left unfinished business when we said that the left-hand side of plot 9 seemed quite straight. We can attack this logarithmic straightness in the early years just as we attacked the linear straightness in the later ones.

Plot 13 shows the result of adding the straight line to plot 8. It is encouraging enough for us to hasten to the residuals, shown in plot 14. As we ought to have expected, they prove useful in the early years but of very dubious value in the later ones. We can again afford to use the magnifying glass on the relatively flat section, as is done in plot 15.

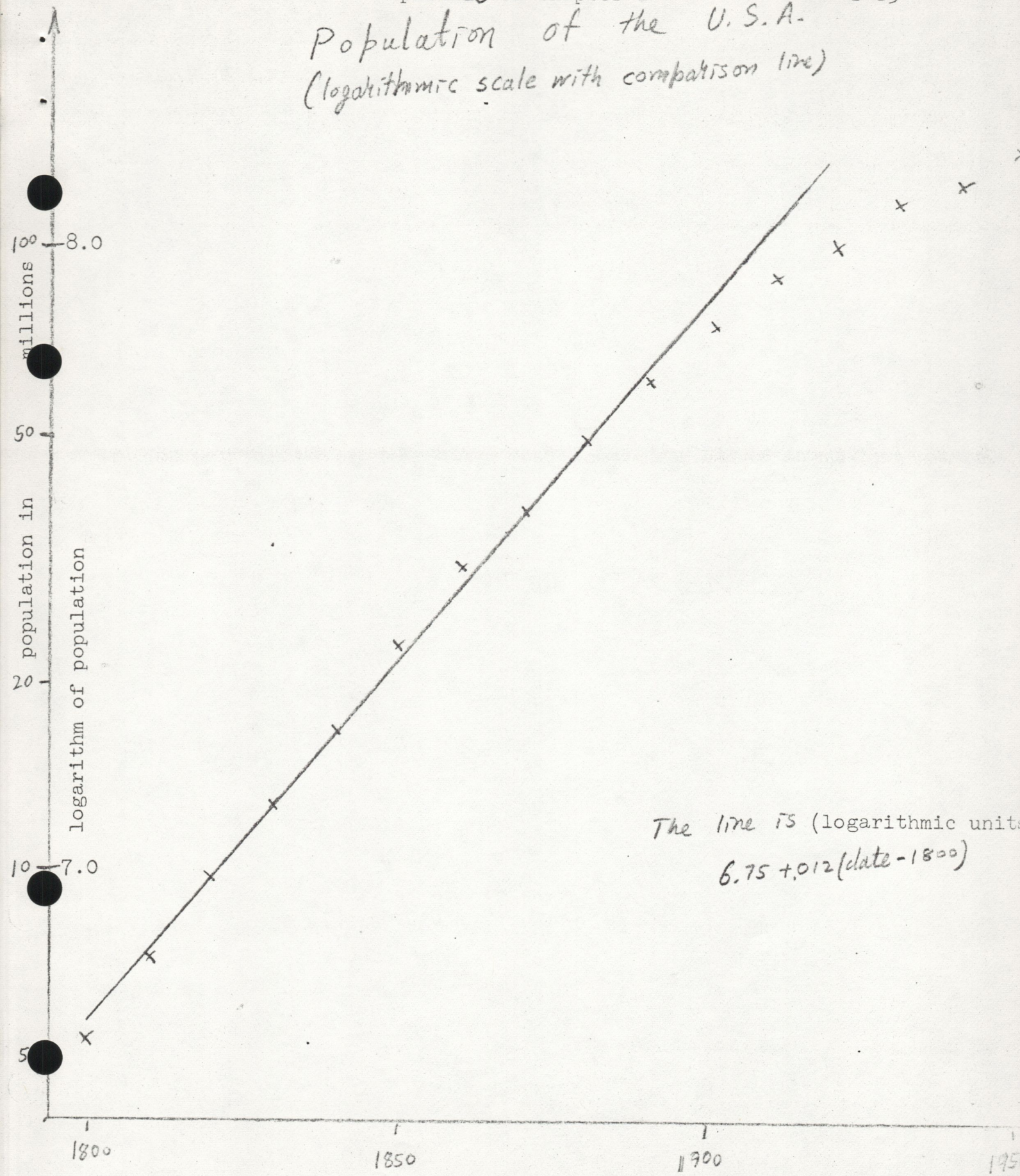
Plot 15, once we realize that ± 0.01 in $\log_{10}$ is about $\pm 2.3\%$ in size, gives us a quite delicate view of U.S.A. population growth during the 19th century. After 1860, population growth was not as fast as before. Moreover, 1810 appears to have been additionally depressed by 3 or 4%. Why? Again we have gone as far with our microscope as seems reasonable without further inputs.

Experts believe many of the fluctuations now so clear are due to variations in completeness of the census, rather than to changes in population growth. Clearly the data do not contradict the experts. The fluctuations are in the numbers. It is worthwhile to find them, whether they tell us about population or censuses.

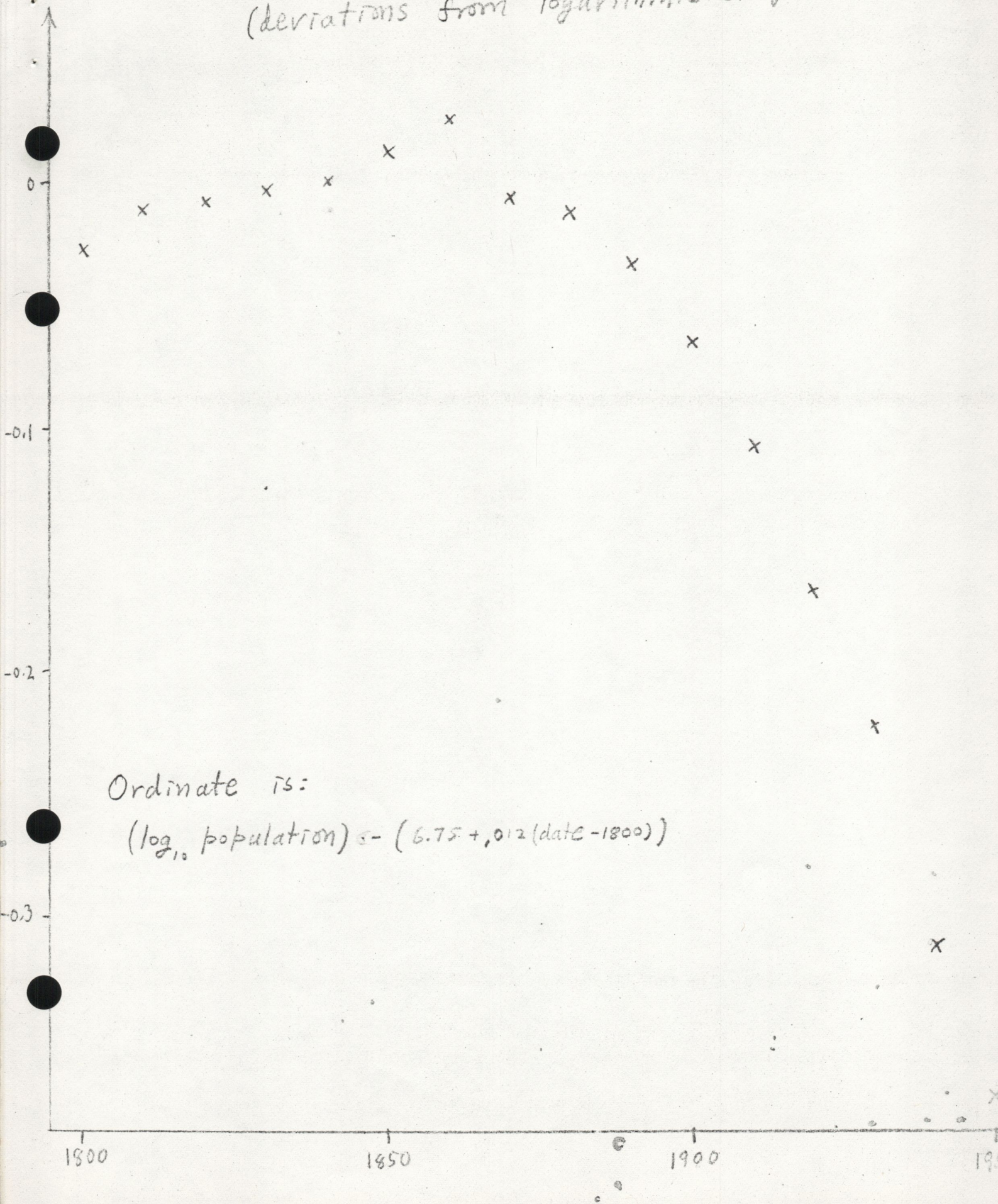
If we had to summarize U.S.A. population growth as completely as we now can in a set of graphs, we would probably choose to show all four of the following:

- plots 13 and 10, to show the general patterns of growth
- plots 15 and 12, to show local behavior.

Population of the U.S.A.
(logarithmic scale with comparison line)



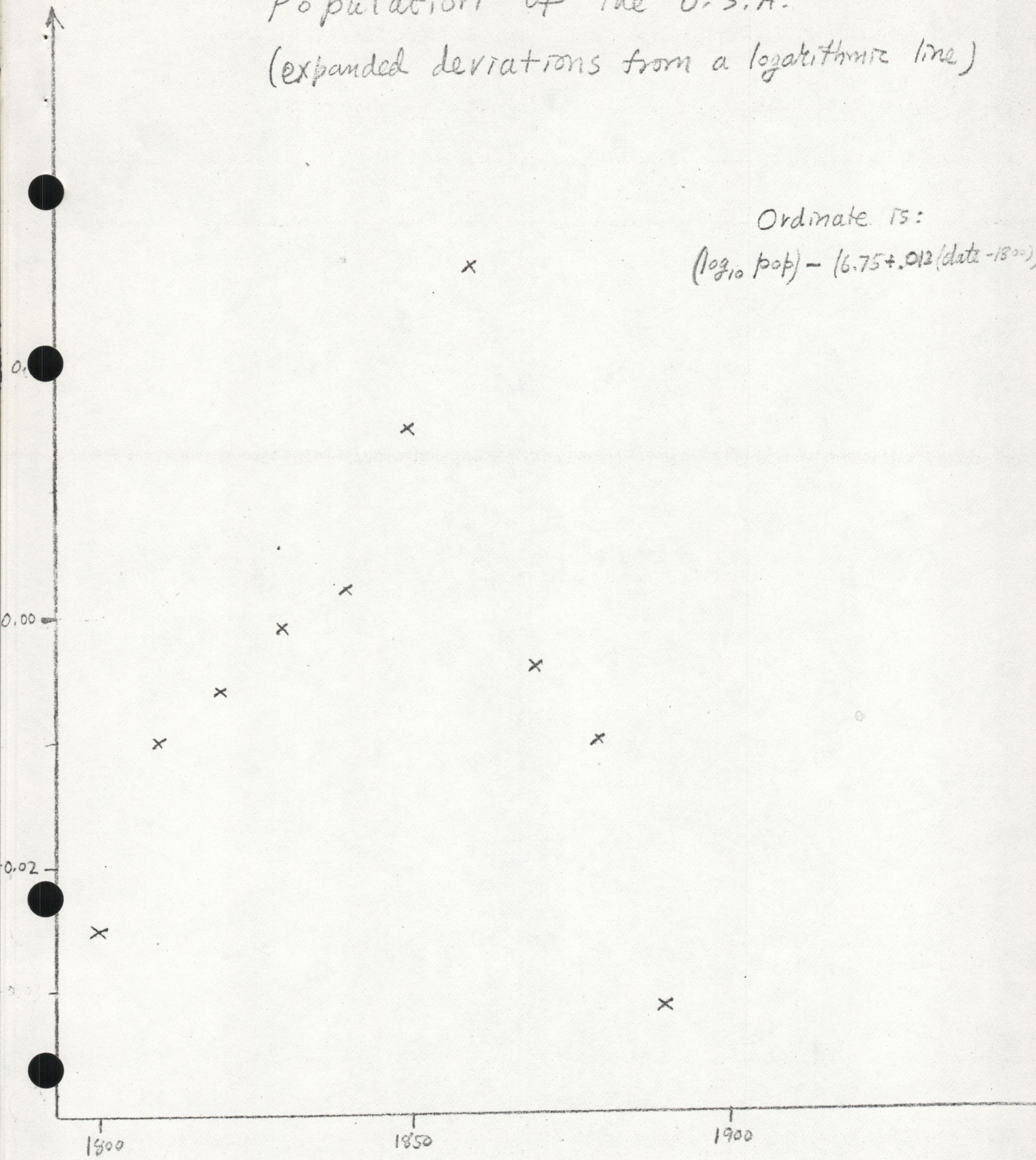
Population of the U.S.A.
(deviations from logarithmic straight line)



Ordinate is:
$(\log_{10} \text{population}) - (6.75 + .012(\text{date} - 1800))$

Population of the U.S.A.

(expanded deviations from a logarithmic line)



(If we were population specialists, we would know about logistic functions and be able to fit a single partial description all the way from 1800 to 1950. This would simplify the plotting of a single graph of residuals, and would probably allow us to summarize the situation in two plots, one showing the fit and the other the residuals.)

What are the lessons to be learned from this example? Not merely that thought can help us see deeper. We have seen specific examples of very general principles, including these:

- choosing scales to make behavior roughly linear always allows us to see local or idiosyncratic behavior much more clearly.
- subtracting partial descriptions to make behavior roughly flat always allows us to expand the vertical scale and look harder at almost any kind of behavior.

Whatever the data, we can try to gain by straightening or flattening. When we succeed in doing one or both, we almost always see more clearly what is going on.

2D. Plotting the ratio of births to deaths

The Census's County and City Data Book contains much varied information. In particular, the 1962 edition gives for each state the number of live births for 1959, the number of deaths for 1959, and the density of population for 1960. One who believed in the "wide-open spaces" might feel that the ratio of births to deaths would appear to be influenced by the density of population, at least if the

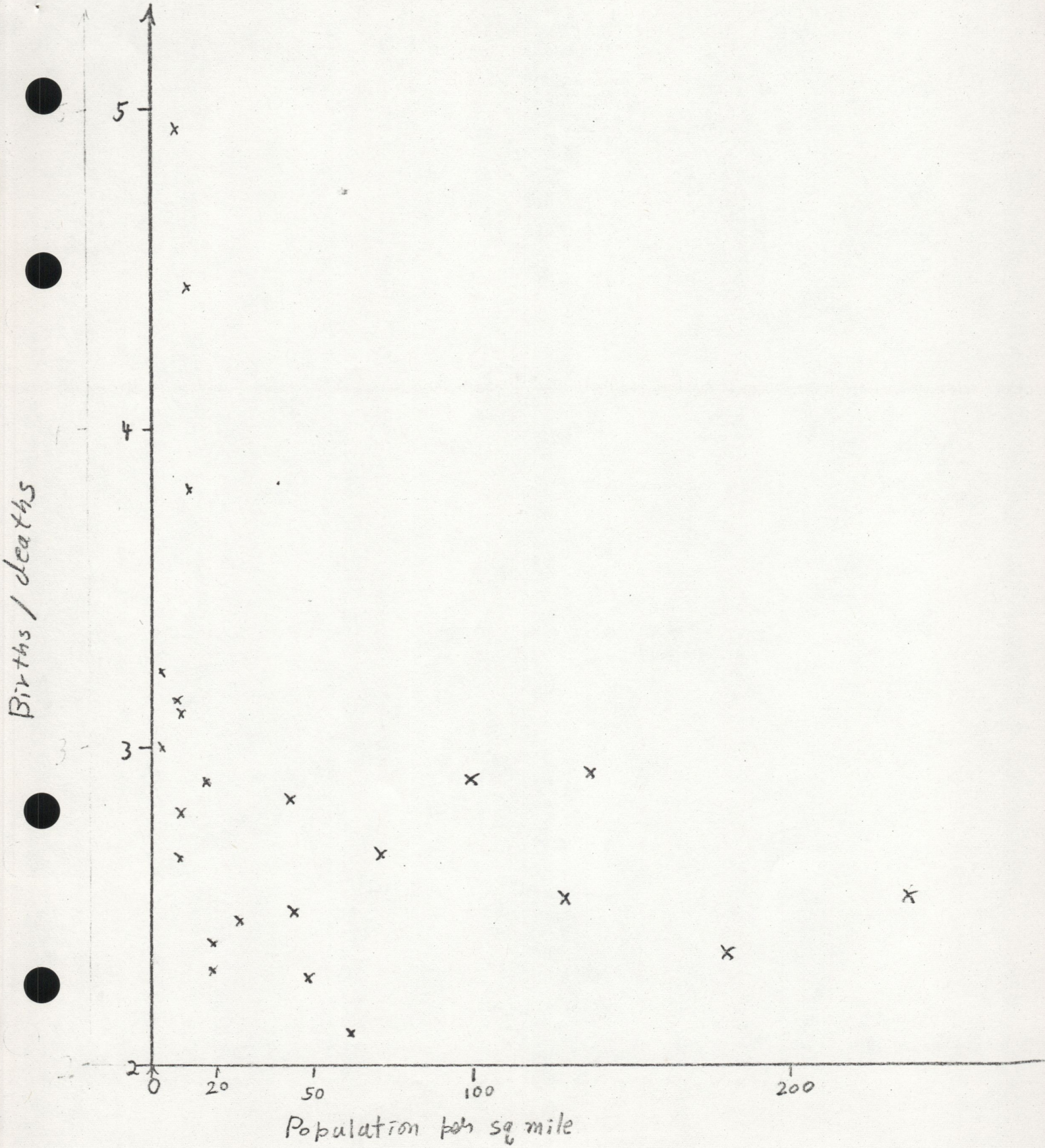
South and the Atlantic coast states were set aside. A plot of the ratio of births to deaths against population density for the remaining states appears as in plot 16. About all we can say from this plot is that the point spatter seems crudely L-shaped. The use of a linear scale for population per square mile has squeezed so many of the states up against the vertical axis that we can't be sure what is going on. If we are to see what, if anything, is going on, we must adjust the left-to-right scale so that the states are less jammed together.

Plot 17 shows the result of using a logarithmic scale for population density. We now see the three states with unusually high ratios of births to deaths have low, but not very low densities, and do not appear to be typical of very low-density states. Looking only at the other states there may be a faint tendency for a higher ratio of births to deaths to go with lower population density. (By omitting the three states, we could double the vertical scale, putting this tendency under more of a microscope. Doing this teaches us almost nothing new, as the reader may verify.)

If we are to give up on population density, what next? The three unusual states (and their birth/death ratios) were: New Mexico (4·95), Utah (4·46), and Arizona (3·83). Looking again in the <u>County and City Data Book</u>, we discover that all three have young populations, for instance as measured by median age. Thus it is natural to plot births/deaths against median age.

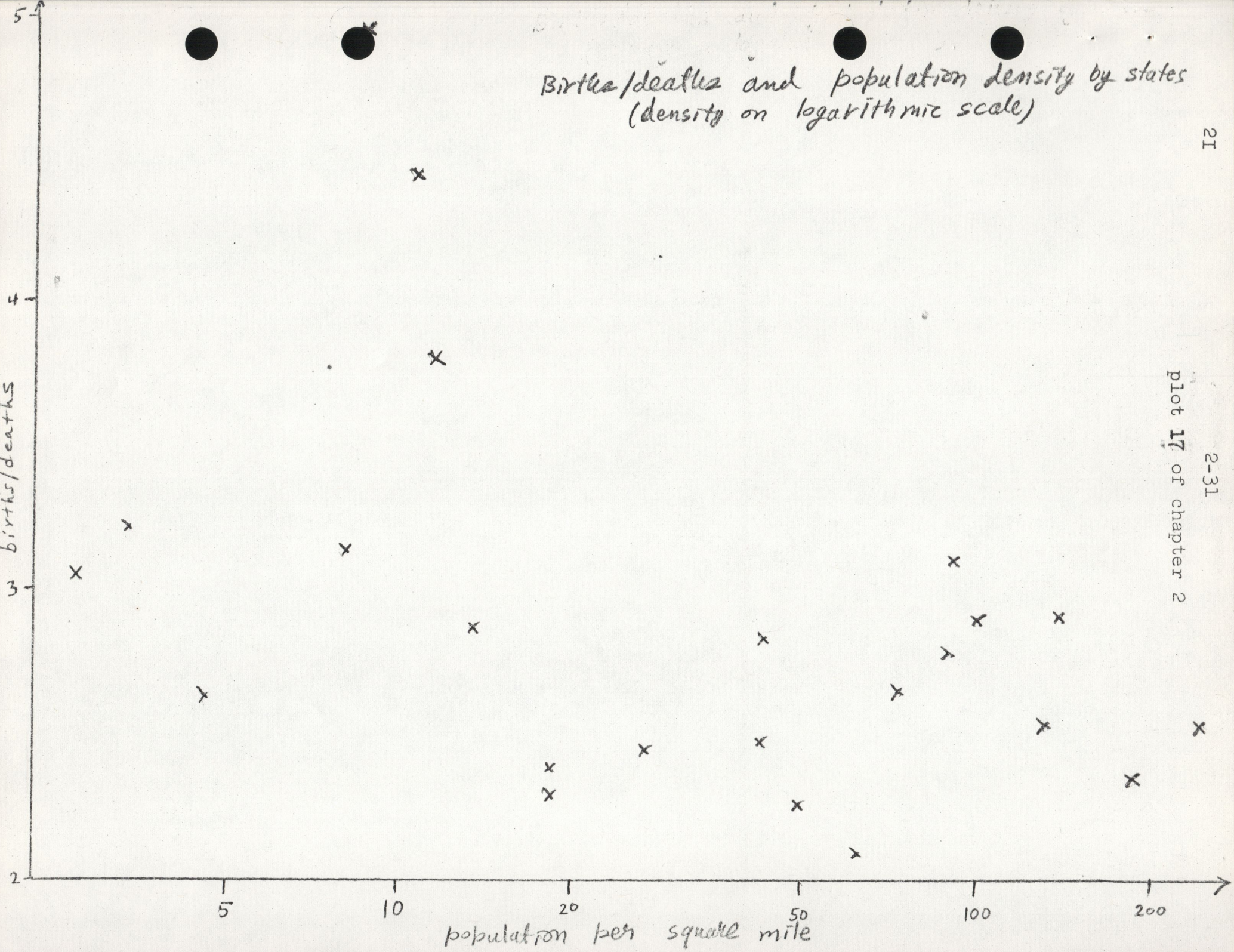
plot 16 of chapter 2

Births/deaths and population density



plot 17 of chapter 2

Births/deaths and population density by states
(density on logarithmic scale)



Plot 18 shows the result. Clearly median age does a much better job of appearing to explain the ratio of births to deaths than did population density, though the apparent dependence is of course not perfect. If we are to go further, we will need to find and subtract out some partial description of this apparent dependence.

When we "put eye to paper" and look along the point spatter in plot 18, we seem to see a definite tendency to curvature (hollow above). If we could subtract this out, we could reasonably compare the individual points with a straight line. How might we approach such a simpler partial description?

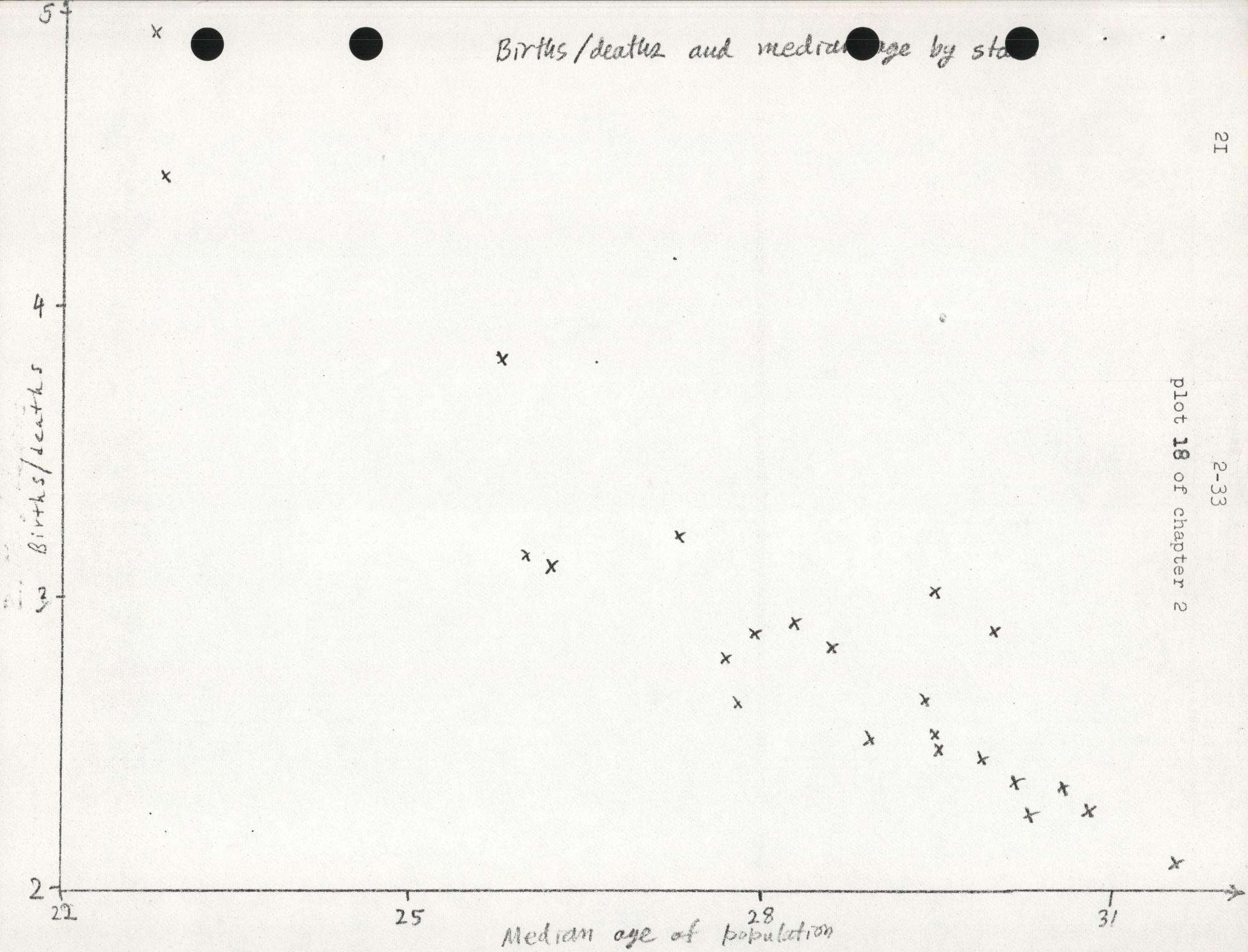
One thing to try is changing the down-to-up scale. It is easy to see that using squares to fix this scale would make curvature worse. So we may as well try going in the opposite direction by using logarithms, a choice which is attractive because of the more symmetric way it treats births and deaths. (Note that

$$\log \frac{\text{births}}{\text{deaths}} = \log \text{births} - \log \text{deaths} = -\log \frac{\text{deaths}}{\text{births}}$$

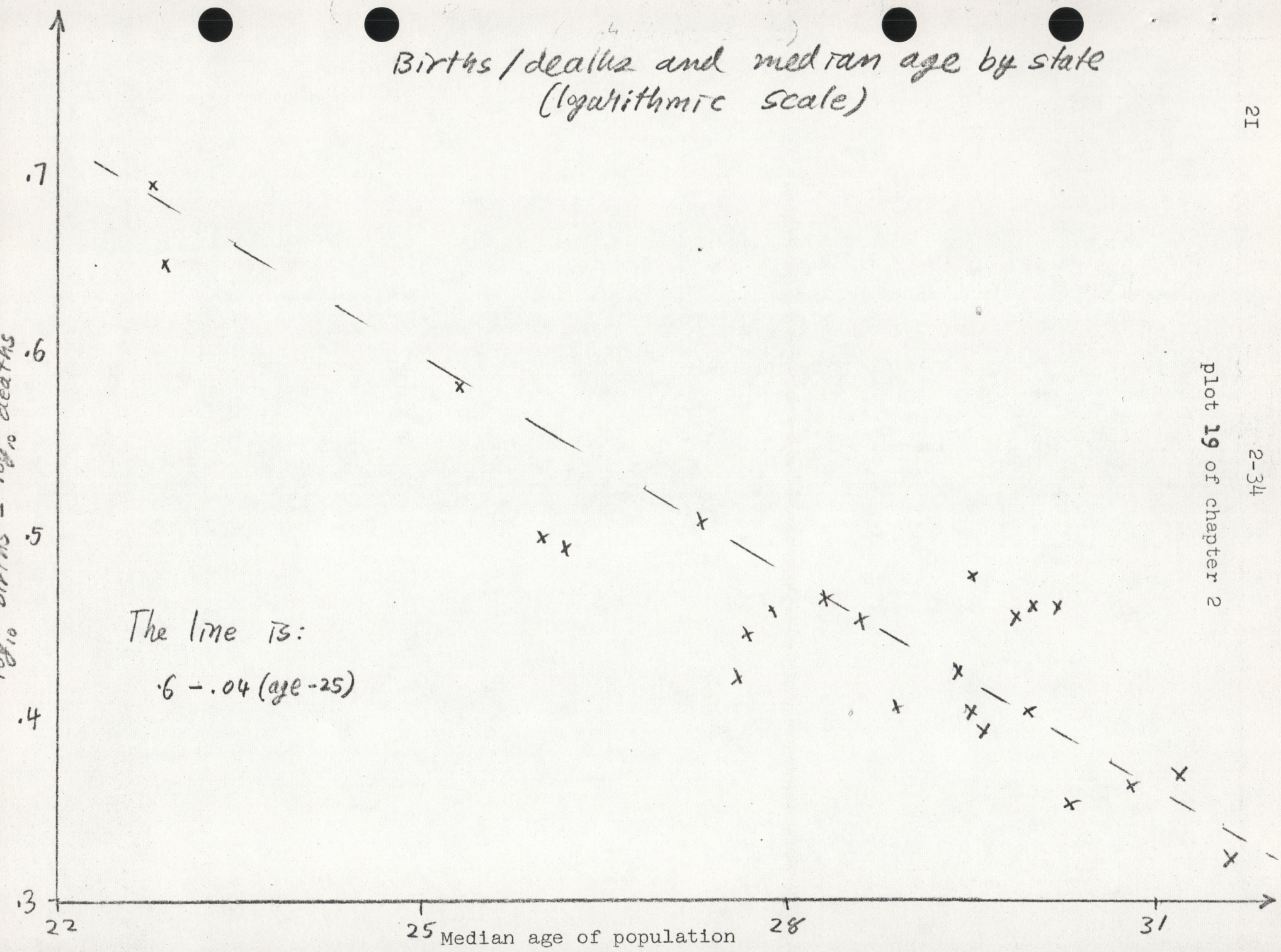
exhibits this symmetry. See Section 2R for background material.)

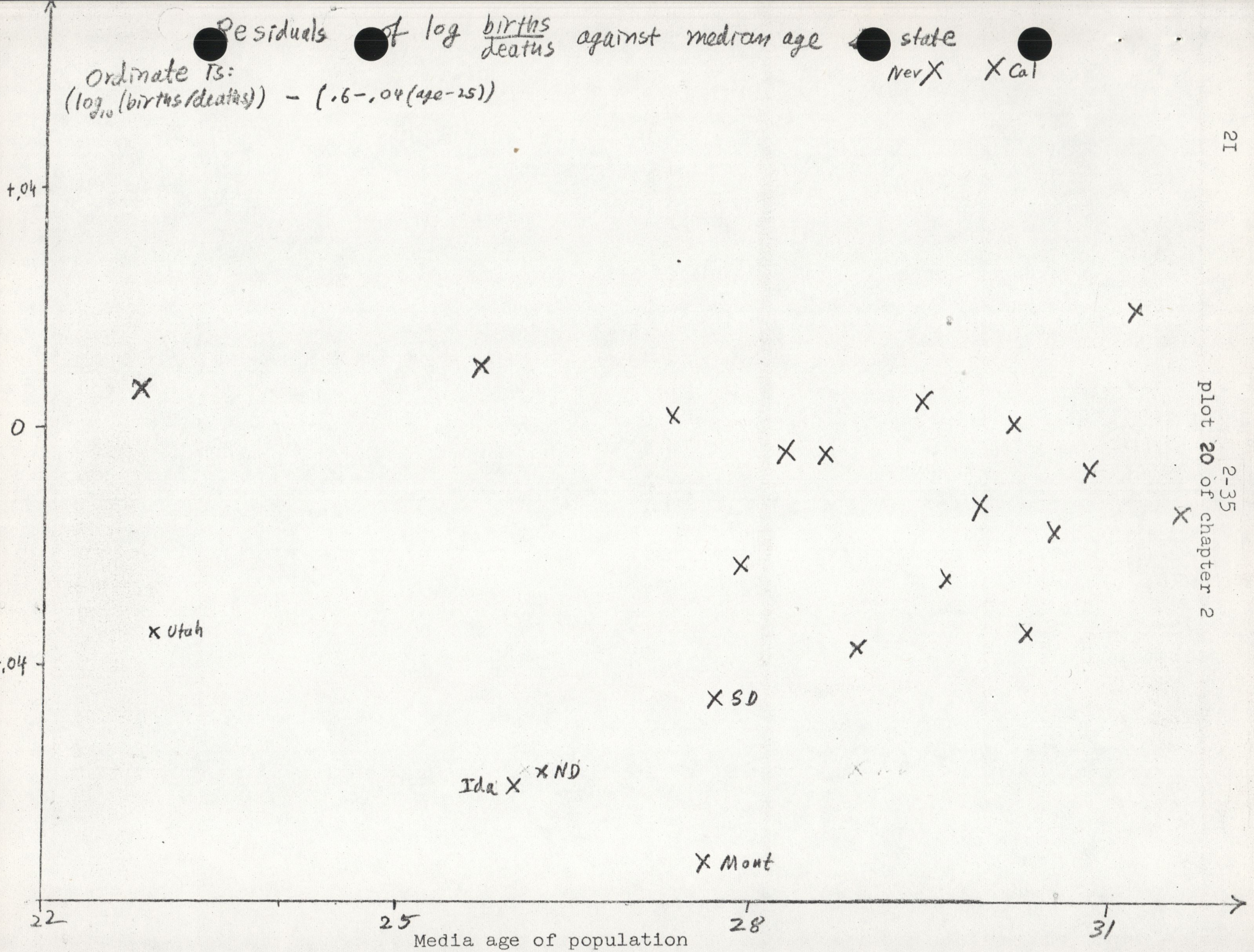
Plot 19 shows the result, complete with a convenient comparison line. The point spatter is now much more nearly straight. When we plot residuals from this line, against median age, and identify the more extreme states, we find the results shown in plot 20.

plot 18 of chapter 2



plot 19 of chapter 2





2-35
plot 20 of chapter 2

The two states with notably high residuals are adjacent to one another on the map. The four states with middle median age and notably low residuals are also contiguous. Clearly we need to see these residuals on a map.

Map 21 shows the residuals spread across a map of the United States. The roughly regular structure of these residuals is rather easily seen -- adjacent states are clearly more often similar than are distant ones. (Might the slightly surprisingly positive residual for Illinois be due to the unusual size of Chicago?) To go further here requires either:

- more careful allowance for the age of the population, or
- more knowledge about the mechanisms affecting birth and death rates in general. (Would state-by-state information on economic conditions help?)

Indeed, both of these are likely to be needed!

In this example, we have again seen the same main points as in the previous one:

- changing scales to make dependences roughly linear usually helps.
- flattening by subtraction makes it much easier to see what is going on at more subtle levels.

The fact that our dependences were approximate rather than exact did not alter these main points at all.

Approximate dependence did, however, bring in one new aspect, as we saw in plots 16 and 17:

- the usefulness of changing scales to reduce confusion caused by crowding.

The residuals of Graph 19 mapped

1 here corresponds to:

± 0.01 in $\log_{10} \frac{\text{births}}{\text{deaths}}$

about $\pm .05$ in $\frac{\text{births}}{\text{deaths}}$

± 0.25 year of median age

Map 21 of chapter 2

2E. How far have we come?

In this chapter we have met graphs. Perhaps we have even come to know them a little.

The state of our progress is not measured by the specific techniques we have seen or understood -- though that kind of progress is essential. Rather our progress is measured by our understanding, and calm acceptance, of such propositions as these:

1) graphs are friendly.

2) arithmetic often exists to make graphs possible.

3) graphs force us to note the unexpected; nothing could be more important.

4) different graphs show us quite different aspects of the same data.

5) there is no more reason to expect one graph to "tell all" than to expect one number to do the same.

6) comparison is not trivial; careful thought helps -- in graphical comparison as well as in arithmetic comparison.

7) "plotting y against x" involves significant choices; how we express one or both variables can be crucial.

8) the first step in penetrating plotting is to straighten out the dependence or point spatter as much as reasonable.

9) plotting y^2, $\sqrt{y}$, log y, -1/y or the like in place of y is one plausible step to take in search of straightness.

10) plotting x^2, $\sqrt{x}$, log x, -1/x or the like instead of x is another.

11) once the plot is straightened, we can usually gain much by flattening it, usually by plotting residuals

(with regard to the partial description implied by the straight line we may not have quite drawn in yet).

12) when plotting point spatters we may need to be careful about how we express x and y in order to avoid concealment by crowding.

13) in looking at a batch for possible separations, graphical techniques can be very valuable.

2K. How to count

We often want to plot information involving many individual values. To do this - as well as to do many other things in data analysis - we need to count, to check off item after item in its proper category or value-class. The classical method of checking off

/
//
///
////
~~////~~
~~////~~ /

could hardly be better designed to promote errors - or slow motion - or both. Few of us, certainly not the writer, can look at /// or //// rapidly and be sure what we see. As a result either we go very slowly, or there are far too many ~~//~~ and ~~////~~ groups, each representing a miscount.

It is easy to do better -- at least in one man's experience. Chart 22 shows how it is easier to count to 10 with sensible marks than to count to 5 the dangerous way. The final result is also easier to turn into numbers. For the writer, and several friends, this way of counting works much better than /, //, ///, ////, ~~////~~. (Some people report initial difficulties, especially in taking ☐ for 10.)

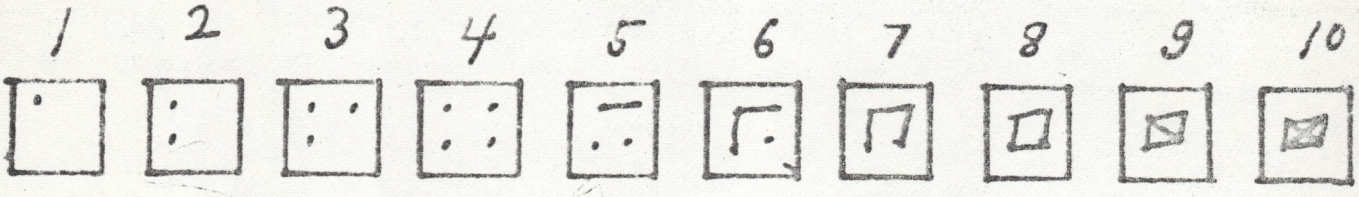
2L. How to plot better

Graphs are important. Great differences in ease of construction and great differences in effectiveness of use

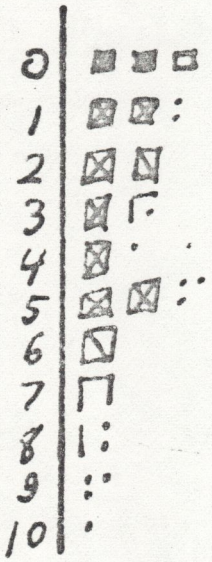
chart 22 of chapter 2

How to count

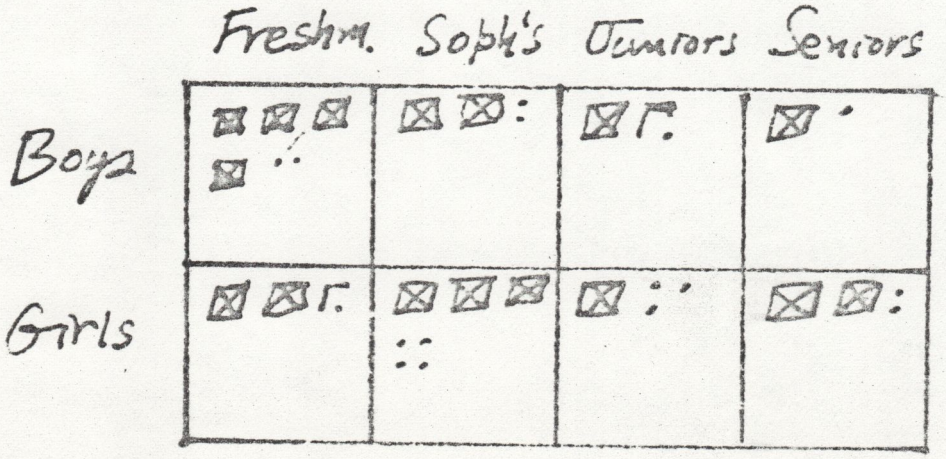
a) The symbols, as built up*



b) A simple example



c) A two-way example



* One need not be consistent, as for example between : and ⁚, unless one wishes to be.

depend on apparently minor, purely technical procedures. This section's remarks are to get you over most of the elementary hurdles.

Choice of ruling. If plotting is to be easy, you want graph paper with at least three different thicknesses of ruling:

- light lines for "units"
- medium lines for "fives"
- heavy lines for "tens"

A wide variety of graph papers can be had with these characteristics. (Some papers have extra heavy lines for "twenties". You will have to learn for yourself whether these bother you more than they help.)

DO NOT USE PAPERS RULED IN "fours" and "eights", OR IN "sixes" AND "twelves" FOR PLOTTING DATA GIVEN IN ORDINARY (decimal) NUMBERS. (For monthly data, of course, "twelves" in one direction do help.)

Avoid quadrille sheets with only one thickness of ruling, and "dime-store" sheets with only two thicknesses if you want to plot fast, easily, and accurately. (If you need to save money, see below.)

Choice of scale units. When you come to plot, you must choose units. Don't try to make one square equal to 3 units, 7 units, 0.03 units, 0.007 units or any such uncomfortable number. Stick to one square being 1, 2, or 5 times a power of ten. (One square = 20,000 or one square = 0.05 are examples that can work out quite well.)

You will find it hard enough to learn to be a fast and accurate plotter with two three-speed gear shifts (1, 2 or 5 across the graph and 1, 2 or 5 up the graph).

If you need to use an abnormal scale - three units to the square, for example - convert your numbers into units of squares - as by dividing by 3 - by slide rule, hand arithmetic, or what have you, BEFORE you start to plot. It will take less time to make a good graph.

Graphs to be looked at, not used to find numbers. Throughout this introduction we shall be interested in graphs to be looked at. Our graphs are means for looking at the data, not stores of quantitative information. This means:

- we will want to keep our eyes on the points,
- we will usually NOT connect one point to the "next" one,
- we will want to suppress the rulings of graph paper, at least from our mind's eye,
- we will need only a few numbered ticks along most axes, horizontal or vertical,
- we will want to use symbols large enough to stand out (and if we need more than one kind, different kinds should be both clearly different and, usually, almost equally noticeable).

There cannot be too much emphasis on our need to see behavior.

We must play down or eliminate anything that might get in the way of our seeing what appears to go on.

Tracing paper. One important tool to this end is the use of tracing paper. If we slip a well-printed sheet of graph paper just below the top sheet of a pad of tracing paper, we can plot on that top sheet of tracing paper almost as easily as if it were itself ruled. Then, when we have the points plotted, some boundary lines drawn, and a few scale points ticked, we can take away the graph sheet and look at the points undisturbed by a grid. We often gain noticeably in insight by doing this. (And we have had to pay for a sheet of tracing paper rather than for a sheet of graph paper.)

Kinds of grids. The use of tracing paper makes it possible for any of us to really arm himself with an easy-graphing capability of broad scope. A tracing pad and one sheet each of 20 kinds of graph paper puts us in business to do many things. (Actually, it usually seems worthwhile to have a reasonable number of sheets, perhaps in the form of a pad, of at least one or two of the most used kinds of paper. Many will wish to first plot on the graph paper and then copy on the tracing paper.)

"Ordinary" graph paper frequently comes 8 to the inch (beware), 10 to the inch, 12 to the inch (beware), 20 to the inch, and 10 to the centimeter. The writer likes the coarser ratings at least as well as the fine; it is up to you to learn your own preference. (Beware of 12 to the inch rulings, since too many are ruled in "sixes" and "twelves"; but bear in mind that 12 to the inch

fits in well with some typewriter spacings, so that 12 to the inch ruled in fives and tens like Codex Nos. 31,253 and 32,253 is a very good base for graphs where the final product is a typed page. It has little special advantage for most other purposes.) Never use 8 to the inch ruled in 4's and 8's (unless what you are plotting comes in 1/8th of an inch).

Semilogarithmic (uniform scale one way, logarithmic scale the other) and full logarithmic (logarithmic scales both ways) graph papers come in a variety of scales and patterns. To have a reasonable selection is to save time and encourage inquiry. One CAN always look up the logarithms and then plot on uniform-scale graph paper, but WILL one? As we saw above, logarithms are often the way to make data reveal itself.

A variety of other graph papers are useful, as we shall see later. We note here that paper with two square-root scales is available (Codex No. 31,298 or No. 32,298) and that trilinear or isometric paper, with three sets of rulings at 120° to one another is often quite useful. (It is available from various manufacturers.) Reciprocal (one way) paper is available from both Codex and K & E.

<u>Shape of plot</u>. Naturally enough, graph paper comes in the same general shape as ruled paper (and writing paper). Clearly it can be used in two ways:

- narrow edge at the top of the plot, or
- broad edge at the top.

By analogy, with the use of ruled paper it is natural to make plots with the narrow edge up. Full use of the area then usually makes a plot that is taller than it is wide.

There are some purposes for which this is a good shape. The ever-faster rising curve of early growth is an example. We can quite well use taller-than-wide plots for such simple pictures, most of which tell us that we haven't yet made an analytically useful plot. Sometimes, indeed, such simple plots can be clearer when they are taller than wide rather than wider than tall.

Most diagnostic plots either involve a more or less definite dependence that bobbles around a lot (like plot 13 above) or a point spatter (like plot 21 above). Such plots are almost always better made wider than tall. Wider-than-tall shapes usually make it easier for the eye to follow from left to right. ("Do as I say, not as I did in plot 13".) For still other purposes, taller-than-wide plots are to be preferred.

When a plot is made wider than tall, convention says it should be turned in the direction illustrated by many plots in this chapter, even if this makes the down-to-up legend upside down when we first see the plot as we turn the pages.

2M. How to plot without graph paper

We almost always want to look at numbers. We do not always have graph paper at hand. This is no excuse for failing to plot and look.

We usually have ruled paper at hand. For emergency graph paper, take out one sheet of ruled paper, turn it on its

side, and place it beneath another sheet of ruled paper. If these two sheets have a light-colored backing -- often provided by the rest of the pad or notebook -- the vertical lines on the lower sheet are almost certain to show through well enough to combine with the horizontal lines on the top sheet to form a grid on which plotting is reasonably easy. (The first step in this sort of plotting is to, by ticks or unobtrusive dots, mark enough information on the top sheet to make it easy to get the lower sheet back to its original position after it slips.)

With this technique, one can make useful, if not decorative, plots almost anywhere.

*****

2P. IEDA processor

*****

2R. How to think about logs

We want to use logarithms -- "logs" -- freely as a tool. The material in this section is designed to make such use easier.

Properties. We recall, for reference, those few important facts about

(1) Logs can be looked up in tables.

(2) Logs come "to" different bases; however logarithms to different bases only differ by multiplicative constants. As a result any base is almost (usually exactly) as good as any other in data analysis.

(3) We shall use "$\log_{10}$" for logs to the base 10 (and would use "$\log_e$" or "ln" for logs to the base e).

(4) The log of 1 is 0.

(5) The log of zero is not defined.

(6) The log of a negative number can be defined, but is not a real number. (We shall not try to use logs of negative numbers.)

(7) The log of a product is the sum of the logs of the factors, thus

$$\log uvw = \log u + \log v + \log w$$

(8) The log of a ratio is the difference of the logs of numerator and denominator, thus:

$$\log(s/t) = \log s - \log t$$

(9) The log to base 10 of 10 is 1, (that of 100 is 2, of 1000 is 3, of 0.1 is -1, of 0.01 is -2, of 0.001 is -3, etc.).

2S. How to get low-precision logs easily

In data analysis, it is frequently quite adequate to calculate logs to low precision. Two decimal places is frequently enough - indeed one decimal will sometimes do quite well. In such situations special tables save effort. To save as much effort as possible, these tables should allow direct entry, with no fuss about interpolation. This can be done by the use of so-called critical tables.

For one-decimal logs we can use the following table.

First nonzero digits of x	First decimal of $\log_{10}x$
8913 to 1122	.0
1123 to 1412	.1
1413 to 1778	.2
1779 to 2238	.3
2239 to 2818	.4
2819 to 3548	.5
3549 to 4466	.6
4467 to 5623	.7
5624 to 7079	.8
7080 to 8912	.9

The plan and working of a critical table is easily seen through by experimentation with this table.

For two-decimal logs we can use table 23, which has been compressed slightly by comparison with the one-decimal example. As a result, we do not know whether to move up or down to find the first decimal when our x exactly matches the break value. Don't worry, do one OR the other. In data analysis, if two decimals are enough, it won't matter how we resolve such questions.

table 23 of chapter 2

Critical Table for Two-Decimal Logs

Break	$\log_{10}$	Break	$\log_{10}$	Break	$\log_{10}$	Break	$\log_{10}$	Break	$\log_{10}$
9886	.00	1567	.20	2483	.40	3936	.60	6237	.80
1012	.01	1603	.21	2541	.41	4027	.61	6383	.81
1035	.02	1641	.22	2600	.42	4121	.62	6531	.82
1059	.03	1679	.23	2661	.43	4217	.63	6683	.83
1084	.04	1718	.24	2723	.44	4315	.64	6839	.84
1109	.05	1758	.25	2786	.45	4416	.65	6998	.85
1135	.06	1799	.26	2851	.46	4519	.66	7161	.86
1161	.07	1841	.27	2917	.47	4624	.67	7328	.87
1189	.08	1884	.28	2985	.48	4732	.68	7499	.88
1216	.09	1928	.29	3055	.49	4842	.69	7674	.89
1245	.10	1972	.30	3126	.50	4955	.70	7852	.90
1274	.11	2018	.31	3199	.51	5070	.71	8035	.91
1303	.12	2065	.32	3273	.52	5188	.72	8222	.92
1334	.13	2113	.33	3350	.53	5309	.73	8414	.93
1365	.14	2163	.34	3428	.54	5433	.74	8610	.94
1396	.15	2213	.35	3508	.55	5559	.75	8810	.95
1429	.16	2265	.36	3589	.56	5689	.76	9016	.96
1462	.17	2317	.37	3673	.57	5821	.77	9226	.97
1496	.18	2371	.38	3758	.58	5957	.78	9441	.98
1531	.19	2427	.39	3846	.59	6095	.79	9661	.99
1567		2483		3936		6237		9908	

1		1	
0.1	-1	10	+0
0.01	-2	100	+1
0.001	-3	1000	+2
0.0001	-4	10,000	+3
0.00001	-5	100,000	+4
0.000001	-6	1,000,000	+5

Examples: $\log 137.2 = 2 + .14 = 2.14$

$\log 0.03694 = -2 + .57 = -1.43$

2T. Critical tables of other handy powers

Since we are likely to want to try $-1/x$ in much the same situations as where we want to try $\log_{10}x$, there is a real reason for putting a critical table of $-1/x$ here also. The table we provide is of precision intermediate between the one-decimal and two-decimal place critical tables of logarithms. It often suffices.

This table is given as table 24. To look up $-1/x$ for $x = 43.7$, we note that 437 comes between 4347 and 4425, so that $-1/x$ starts with -228 (to this accuracy). Fixing the decimal point sensibly gives us -.0228 for $-1/x$.

We also want to make it easy to use $\sqrt{x}$ and $-(1/\sqrt{x})$. Accordingly tables 25 and 26 give low precision values for these.

table 24 of chapter 2

Critical Table for -1/x

Break	-1/x	Break	-1/x	Break	-1/x	Break	-1/x	Break	-1/x
990		1639		2469		4115		610	
1010	-100	1681	-60	2532	-40	4202	-240	633	-160
1030	-98	1709	-59	2597	-39	4274	-236	649	-156
1053	-96	1739	-58	2667	-38	4347	-232	666	-152
1075	-94	1770	-57	2740	-37	4425	-228	685	-148
1099	-92	1802	-56	2816	-36	4504	-224	704	-144
1124	-90	1835	-55	2899	-35	4587	-220	725	-140
1149	-88	1869	-54	2985	-34	4672	-216	746	-136
1176	-86	1905	-53	3077	-33	4762	-212	769	-132
1205	-84	1942	-52	3175	-32	4854	-208	793	-128
1235	-82	1980	-51	3287	-31	4950	-204	820	-124
1266	-80	2020	-50	3367	-30	505	-200	840	-120
1299	-78	2062	-49	3448	-294	515	-196	855	-118
1333	-76	2105	-48	3509	-288	526	-192	870	-116
1370	-74	2151	-47	3584	-282	538	-188	885	-114
1408	-72	2198	-46	3663	-276	549	-184	901	-112
1449	-70	2247	-45	3745	-270	562	-180	917	-110
1493	-68	2299	-44	3831	-264	575	-176	935	-108
1538	-66	2353	-43	3922	-258	588	-172	952	-106
1587	-64	2410	-42	4016	-252	610	-168	971	-104
1639	-62	2469	-41	4115	-246	633	-164	990	-102

Note. Watch decimal places!

Examples: -1/124.2 = -.0080

-1/.04739 = -21.2

table 25 of chapter 2

Critical Table for $\sqrt{x}$

Break	$\sqrt{x}$	Break	$\sqrt{x}$	Break	$\sqrt{x}$	Break	$\sqrt{x}$	Break	$\sqrt{x}$
98 01	100	2 49 64	160	5 66	240	15 60	40	35 40	60
1 02 01	102	2 62 44	164	5 86	246	16 40	41	31 21	62
1 06 09	104	2 75 56	168	6 20	252	17 22	42	39 69	64
1 10 25	106	2 89 00	172	6 50	258	18 06	43	42 25	66
1 14 49	108	3 02 76	176	6 81	264	18 92	44	44 69	68
1 18 81	110	3 16 84	180	7 13	270	19 80	45	47 61	70
1 23 21	112	3 31 24	184	7 45	276	20 70	46	50 41	72
1 27 69	114	3 45 96	188	7 78	282	21 62	47	53 29	74
1 32 25	116	3 61 00	192	8 12	288	22 56	48	56 25	76
1 36 89	118	3 76 36	196	8 47	294	23 52	49	59 79	78
1 41 61	120	3 92 04	200	8 82	30	24 50	50	62 41	80
1 48 84	124	4 08 04	204	9 30	31	25 50	51	65 61	82
1 58 76	128	4 24 36	208	9 92	32	26 52	52	68 89	84
1 69 00	132	4 41 00	212	10 56	33	27 56	53	72 25	86
1 79 56	136	4 57 96	216	11 22	34	28 62	54	75 69	88
1 90 44	140	4 75 24	220	11 90	35	29 70	55	79 21	90
2 01 64	144	4 92 84	224	12 60	36	30 80	56	82 81	92
2 13 16	148	5 10 76	228	13 32	37	31 92	57	86 49	94
2 25 00	152	5 29 00	232	14 06	38	33 06	58	90 25	96
2 37 16	156	5 47 56	236	14 82	39	34 22	59	94 09	98
2 49 64		5 66 44		15 60		35 40		98 01	

Examples: 124.2 = 1 24.2, hence $\sqrt{124.2} = 11.2$

1242 = 12 42. , hence $\sqrt{1242} = 35$

2447.59 = 24 47.59, hence $\sqrt{2447.59} = 49$

table 26 of chapter 2

Critical Table for - $(1/\sqrt{x})$

Break	$\frac{-1}{\sqrt{x}}$	Break	$\frac{-1}{\sqrt{x}}$	Break	$\frac{-1}{\sqrt{x}}$	Break	$\frac{-1}{\sqrt{x}}$	Break	$\frac{-1}{\sqrt{x}}$
98 0	-100	2 68 7	-60	6 10	-40	17 08	-240	38 1	-160
1 02 0	-98	2 82 5	-59	6 41	-39	17 65	-236	40 1	-156
1 06 3	-96	2 92 2	-58	6 75	-38	18 26	-232	42 2	-152
1 10 8	-94	3 02 5	-57	7 11	-37	18 90	-228	44 9	-148
1 15 6	-92	3 13 3	-56	7 51	-36	19 58	-224	46 9	-144
1 20 8	-90	3 24 6	-55	7 93	-35	20 29	-220	49 7	-140
1 26 2	-88	3 36 7	-54	8 41	-34	21 04	-216	52 5	-136
1 32 1	-86	3 49 4	-53	8 91	-33	21 84	-212	55 7	-132
1 38 4	-84	3 62 6	-52	9 47	-32	22 68	-208	59 2	-128
1 45 2	-82	3 77 0	-51	10 08	-31	23 56	-204	63 0	-124
1 52 4	-80	3 92 1	-50	10 75	-30	24 51	-200	67 2	-120
1 60 2	-78	4 08 1	-49	11 34	-294	25 51	-196	70 6	-118
1 68 7	-76	4 25 1	-48	11 81	-288	26 57	-192	73 1	-116
1 77 8	-74	4 43 2	-47	12 33	-282	27 70	-188	75 6	-114
1 87 7	-72	4 62 5	-46	12 85	-276	28 91	-184	78 3	-112
1 98 4	-70	4 83 0	-45	13 42	-270	30 19	-180	81 2	-110
2 10 0	-68	5 05 0	-44	14 63	-264	31 56	-176	84 2	-108
2 22 8	-66	5 28 5	-43	14 68	-258	33 03	-172	87 3	-106
2 36 7	-64	5 53 6	-42	15 38	-252	34 60	-168	90 7	-104
2 52 0	-62	5 80 6	-41	16 13	-246	36 29	-164	94 3	-102
2 68 7		6 09 7		17 08		38 10		98 0	

Examples: 124.2 = 1 24.2, hence $-1/\sqrt{124.2} = -0.090$

.001242 = .00 12 42, hence $-1\sqrt{.001242} = -28.8$

2X. Examples

a) The vapor pressure of H_2O

The vapor pressure of water or ice is the highest pressure of water vapor (steam, if you like) that can exist in equilibrium at a given temperature. Its values are well known, and of considerable practical importance. In terms of temperatures Celsius (once called centigrade, giving 0°C = 32°F and 100°C = 212°F) and pressures in mm of mercury (760 mm is one standard atmosphere) these vapor pressures at different temperatures are given in table 27.

Just a look at the first two columns of this table shows that a direct plot of p against t will only show us a very rapidly rising pressure. (If 139893.20 can be plotted, even 760.00 can hardly be seen.) Accordingly we plan to do something to at least partly straighten out this plot. Taking the logarithm of p looks as if it might help.

Column (3) of table 27 contains the values of $\log_{10}p$. This looks conceivably plottable, so we proceed to plot 28, which can be looked over, but is far from straight.

Another step needs to be taken. We could try to find some empirical approach, but we can go more directly to our goal if we recall that plotting log pressure against the reciprocal of the absolute temperature is rather common in physical chemistry. Here the absolute temperature is, for example, the Celsius temperature increased by about 273.1°C.

Column (4) of table 27 contains values of -1000(1/T), where T = t + 273.1°C. (The factor 1000 has been slipped in to avoid unsightly zeros.) The result of plotting log p

table 27 of chapter 2

The Vapor Pressure of H_2O and Some Associated Quantities

t(°C)	p(mm Hg)	(3)	(4)	(5)
-40	0.105	-.9788	-4.2900	8.6737
-20	0.787	-.1040	-3.9510	8.7858
0	4.5687	.6598	-3.6617	8.8986
20	17.363	1.2396	-3.4118	8.9162
40	54.865	1.7393	-3.1939	8.9256
60	148.88	2.1729	-3.0021	8.9276
80	354.87	2.5501	-2.8321	8.9223
100	760.00	2.8808	-2.6802	8.9112
120	1489.14	3.1729	-2.5439	8.8966
140	2710.92	3.4331	-2.4207	8.8797
160	4636.00	3.6661	-2.3089	8.8611
180	7520.20	3.8762	-2.2070	8.8420
200	11659.16	4.0667	-2.1137	8.8225
220	17395.64	4.2404	-2.0280	8.8034
240	25100.52	4.3997	-1.9489	8.7847
260	35188.00	4.5464	-1.8758	8.7670
280	48104.20	4.6822	-1.8080	8.7502
300	64432.80	4.8091	-1.7449	8.7351
320	84686.80	4.9278	-1.6861	8.7215
340	109592.00	5.0395	-1.6311	8.7095
360	139893.20	5.1458	-1.5795	8.6997

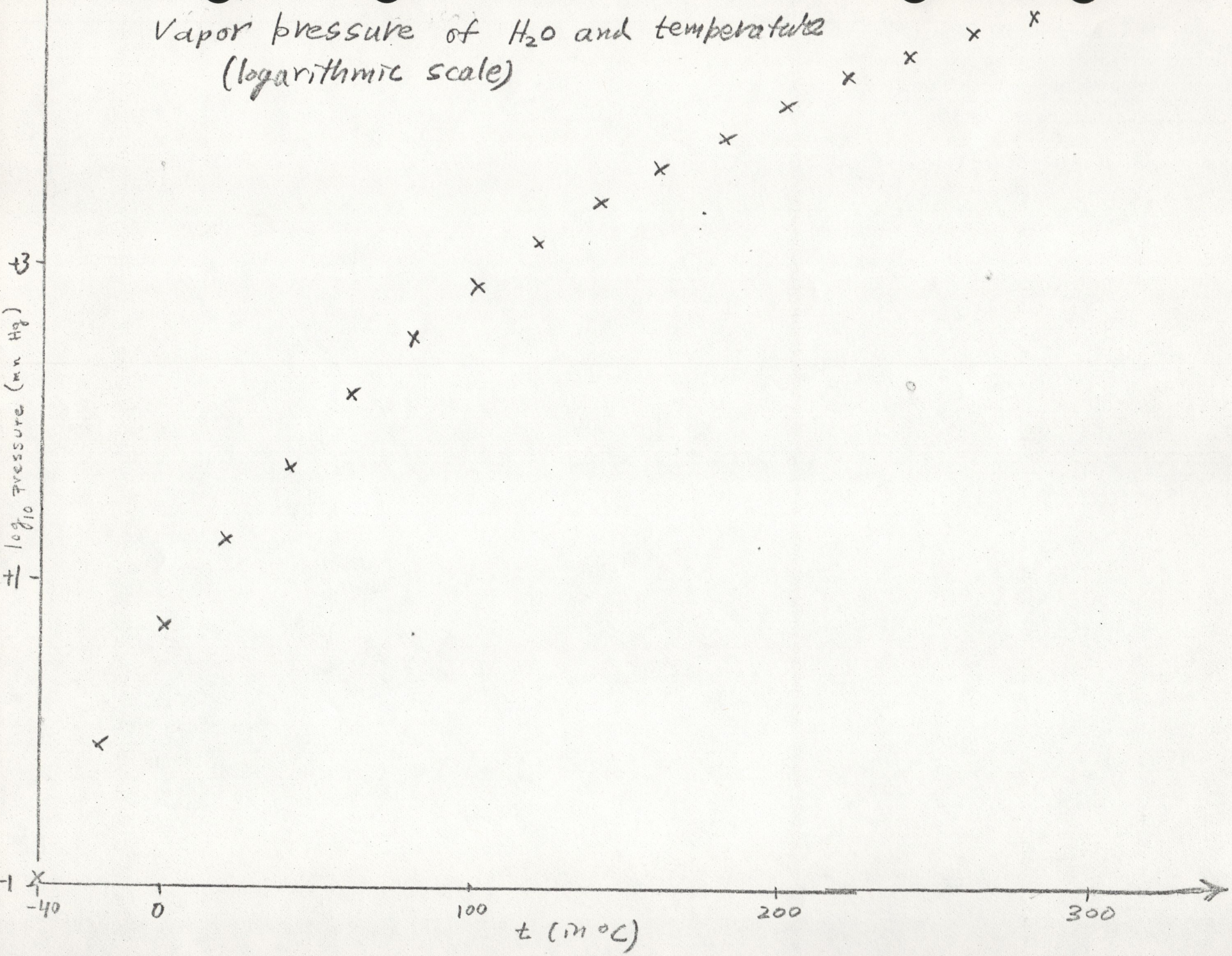
(3) = $\log_{10}p$

(4) = $-1000(1/T)$, where $T = t + 273.1°C$

(5) = $\log_{10}p - 0.225(-1000/T)$

plot 28 of chapter 2

Vapor pressure of H_2O and temperature
(logarithmic scale)



against column (3) of table 27 is shown in plot 29. At last we have a reasonably good straight line, one worth calculating residuals about.

Column (5) of table 27 contains values of $\log p - 0.225(-1000/T)$ which, as we see, is reasonably constant. Subtracting 8.8 clearly gives residuals well worth plotting. Plot 30 shows the result. The most striking result is the apparent split of our curve into the parts with a corner at the third point from the left. What might this mean?

When we recall that this point comes at 0°C (= 32°F) which is the freezing point of water, we see that a break at this point is quite reasonable. Below 0°C we are dealing with the vapor pressure of solid water (ice), while above this point we are dealing with the vapor pressure of liquid water!

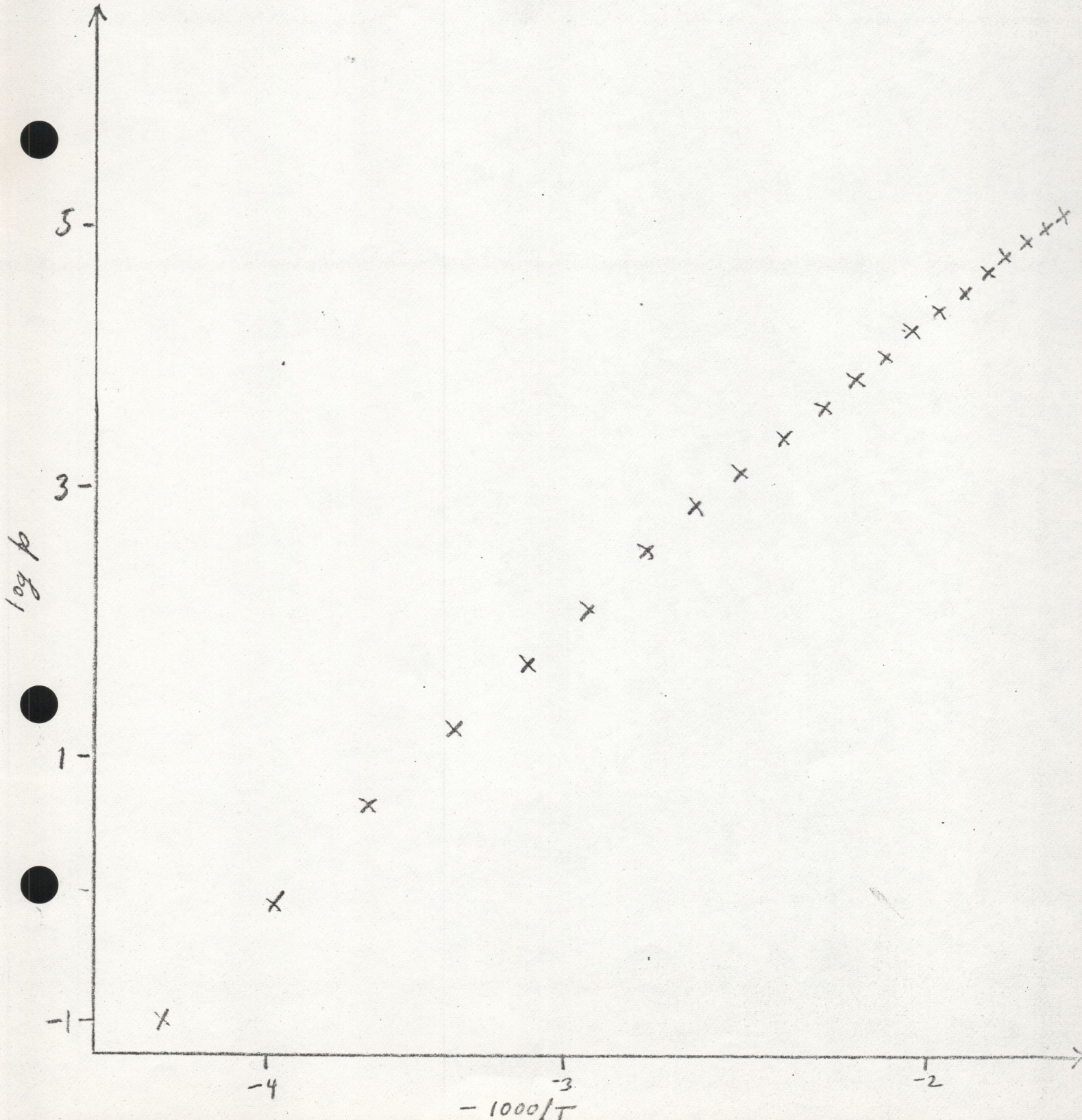
Note that this fact was not forced upon us until we flattened the plot.

b) <u>The decay of radioactivity</u> ---

In 1905, the study of radioactive substances was in its infancy. Meyer and von Schweidler reported the relative activities for an experimental object set out in table 31. If we plot activity against time, the result is plot 32 which is far from being straight. It is again natural to take logarithms, especially since, in simple radioactive decay, the logarithm of activity decreases linearly with time.

plot 29 of chapter 2

Vapor pressure of H_2O and temperature
(logarithm of pressure, reciprocal of absolute temperature)



plot 30 of chapter 2

Residuals of log vapor pressure from a straight line in $-1/T$.
(water)

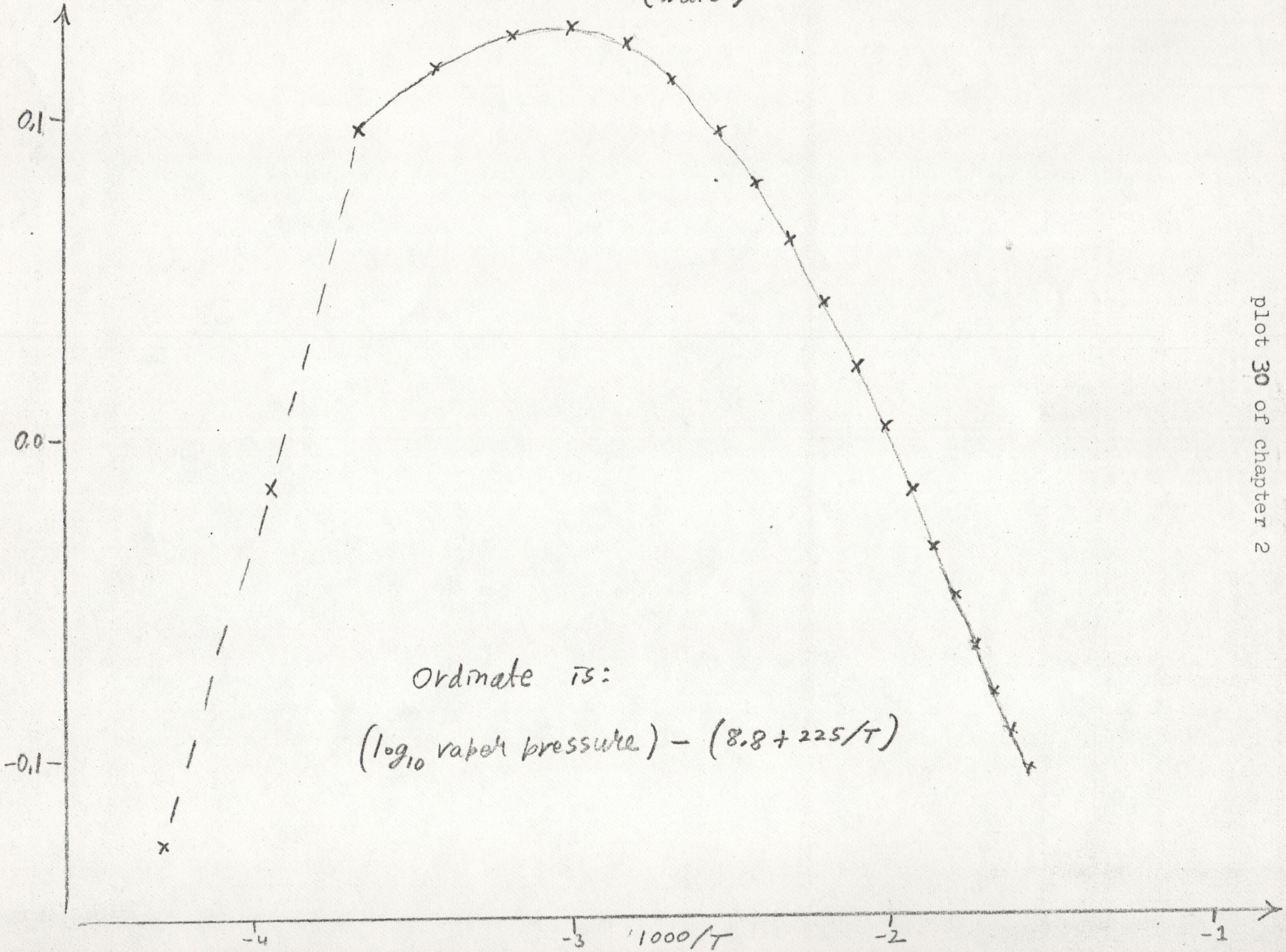


table 31 of chapter 2

The Decay of Radioactivity and Associated Quantities

time (days)	activity (relative)	(3)	(4)	(5)	(6)	(7)
0.2	36.0	1.556	1.544	1.538	1.551	1.545
2.2	26.0	1.415	1.398	1.389	1.471	1.466
4.0	23.1	1.364	1.344	1.334	1.473	1.474
5	18.9	1.276	1.253	1.241	1.418	1.416
6	17.8	1.250	1.225	1.212	1.421	1.422
8	14.7	1.167	1.137	1.121	1.401	1.401
11	13.4	1.127	1.093	1.076	1.456	1.461
12	11.3	1.053	1.013	.991	1.409	1.411
15	8.5	.929	.875	.845	1.370	1.370
18	5.9	.771	.690	.643	1.286	1.273
26	5.0	.699	.602	.544	1.460	1.454
33	3.4	.534	.380	.279	1.469	1.434
39	2.4	.380	.146	-.046	1.431	1.319
45	2.1	.322	.041	-.222	1.526	1.353

(4) = $\log_{10}$(activity - 1.0)

(5) = $\log_{10}$(activity - 1.5)

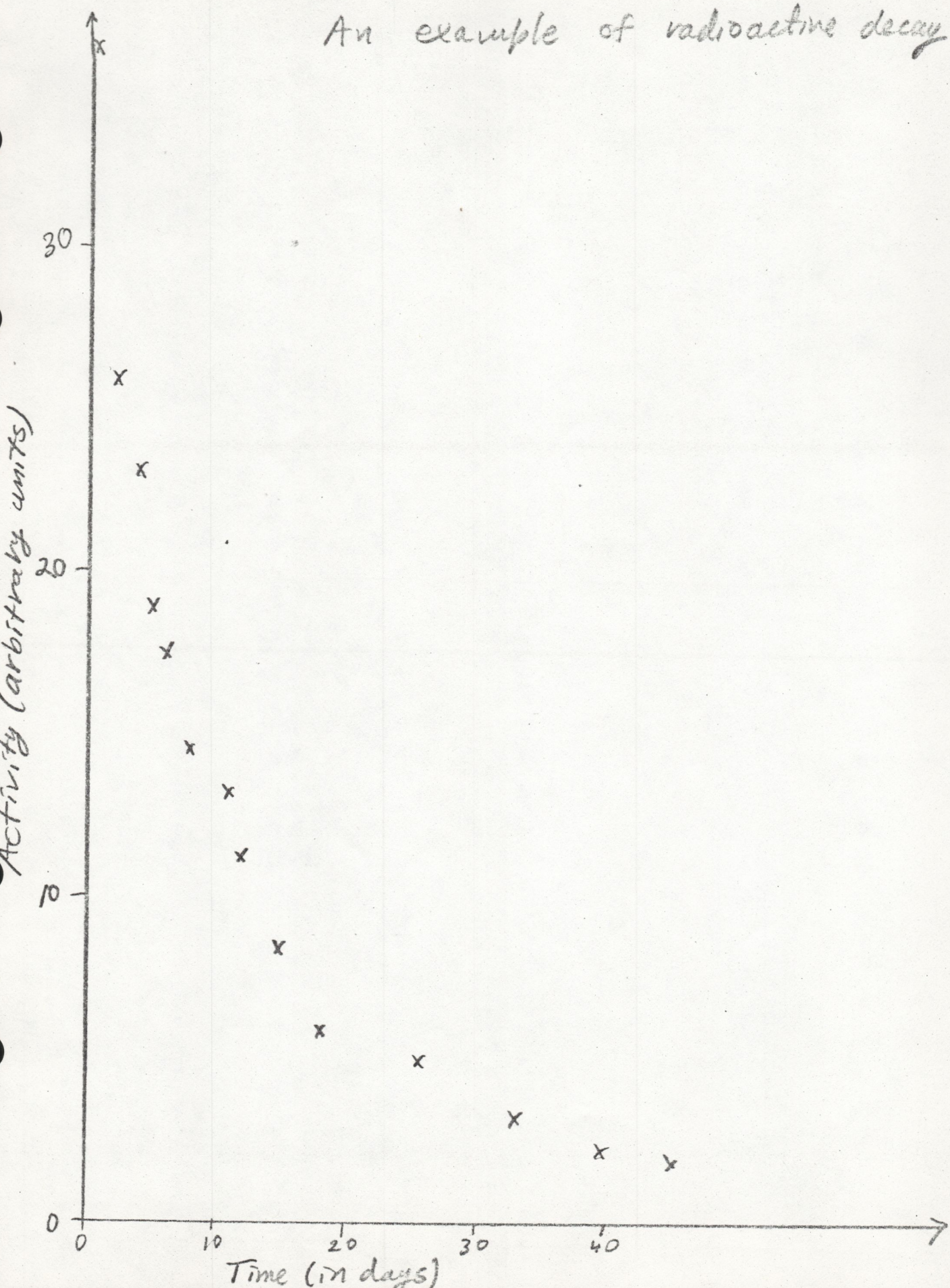
(6) = .033t + (column 4)

(7) = .033t + (column 5)

S. Meyer and E. von Schweidler 1905. Sitzungsberichte der Akademie der Wissenschaften zu Wien, Mathematisch-Naturwissenschaftliche Classe, p. 1202 (Table 5).

plot 32 of chapter 2

An example of radioactive decay



Plot 33 shows the behavior of log activity against time. A noticeable curvature remains. In this situation the most plausible source of curvature is contamination by some other radioactive substance that decays much more slowly than the one of central interest. If such there be, it will not have contributed more than two units of activity (since at 45 days the observed activity is down to 211 units).

It is reasonable to explore the consequences of assuming that 1.0 to 1.5 units of activity from such a contaminant is present by examining the results of forming:

$$\log_{10}(\text{activity} - 1.0)$$

$$\log_{10}(\text{activity} - 1.5)$$

whose values are given in columns (4) and (5) of table 31, fitting a rough comparison line and plotting the residuals. The results are shown in plot 34. (Columns (6) and (7) of table 31 give these residuals increased by 1.45, calculated as ".033t + (column 4)" and ".035t + (column 5)" respectively.) Of the two choices, allowance for 1.0 unit of contamination seems to be a closer fit.

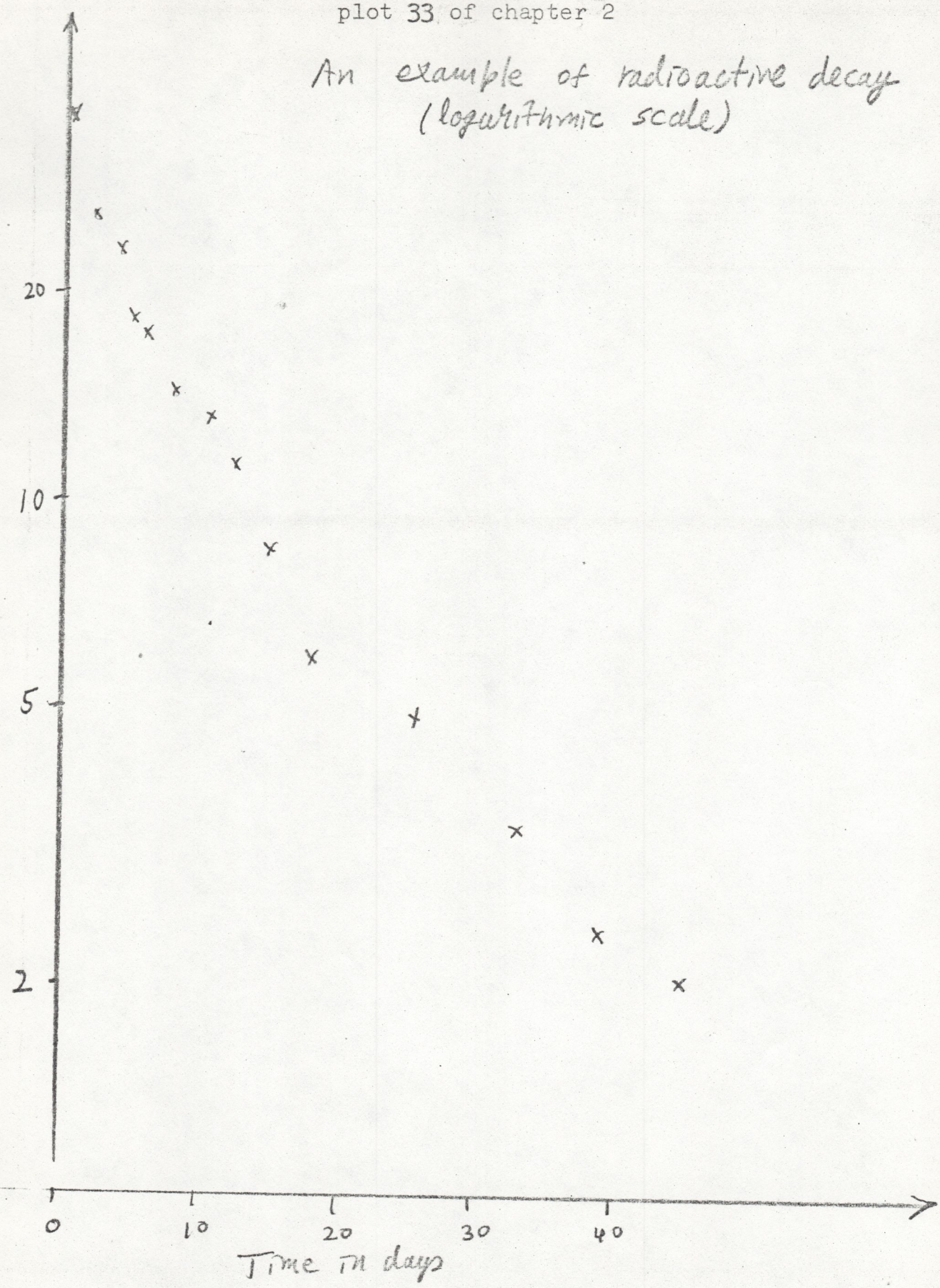
2Y. Problems

Problems for Hand Plotting (and Arithmetic)

1. Tables 35 to 38 contain selected information about the 15 smallest counties in North Carolina, South Carolina, Alabama and Georgia. (1962 City and County Data Book.) Plot y = % with less than five years schooling against x = % with less than $3000 income for at least two states. Continue the analysis. Comment.

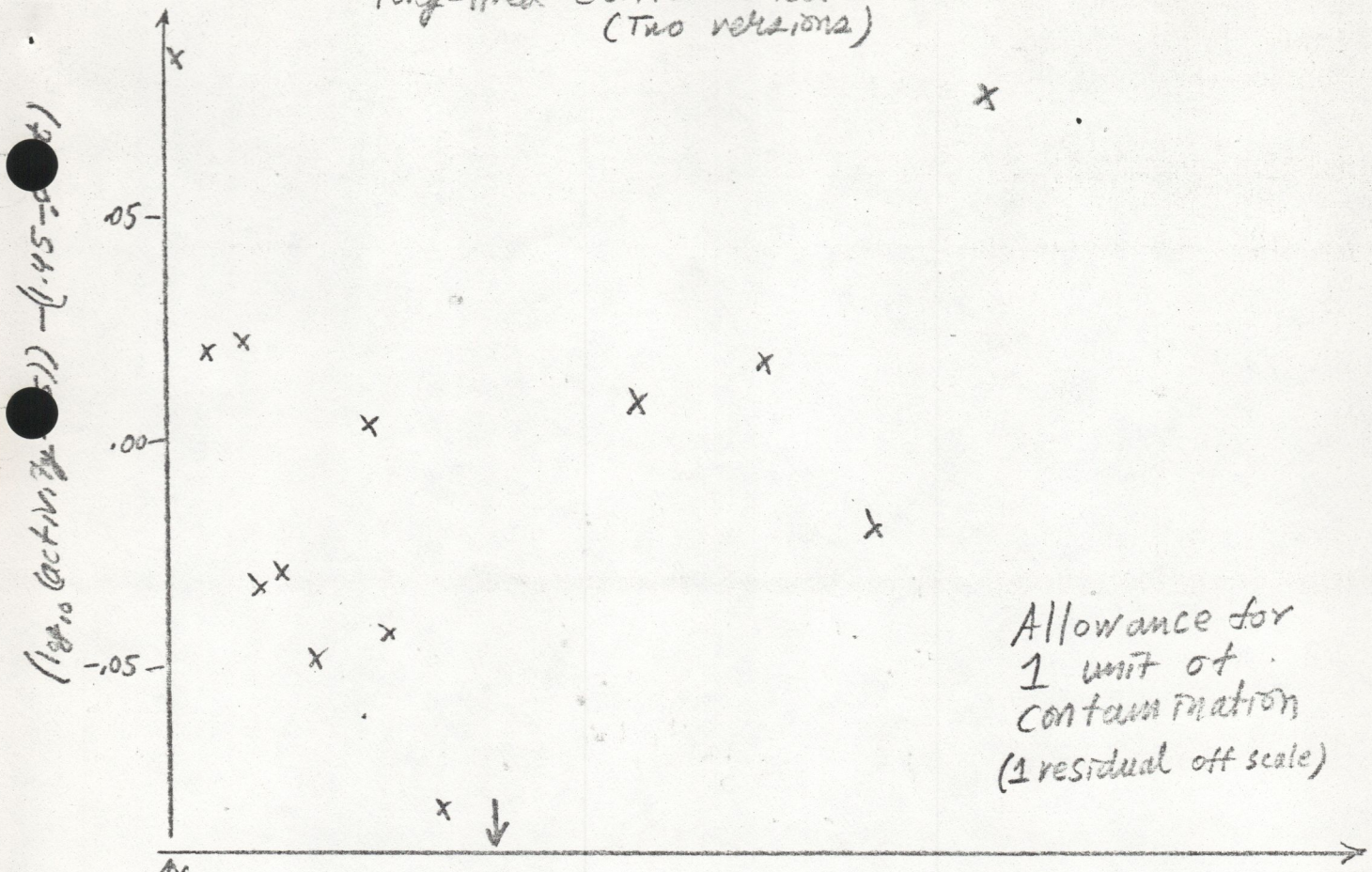
plot 33 of chapter 2

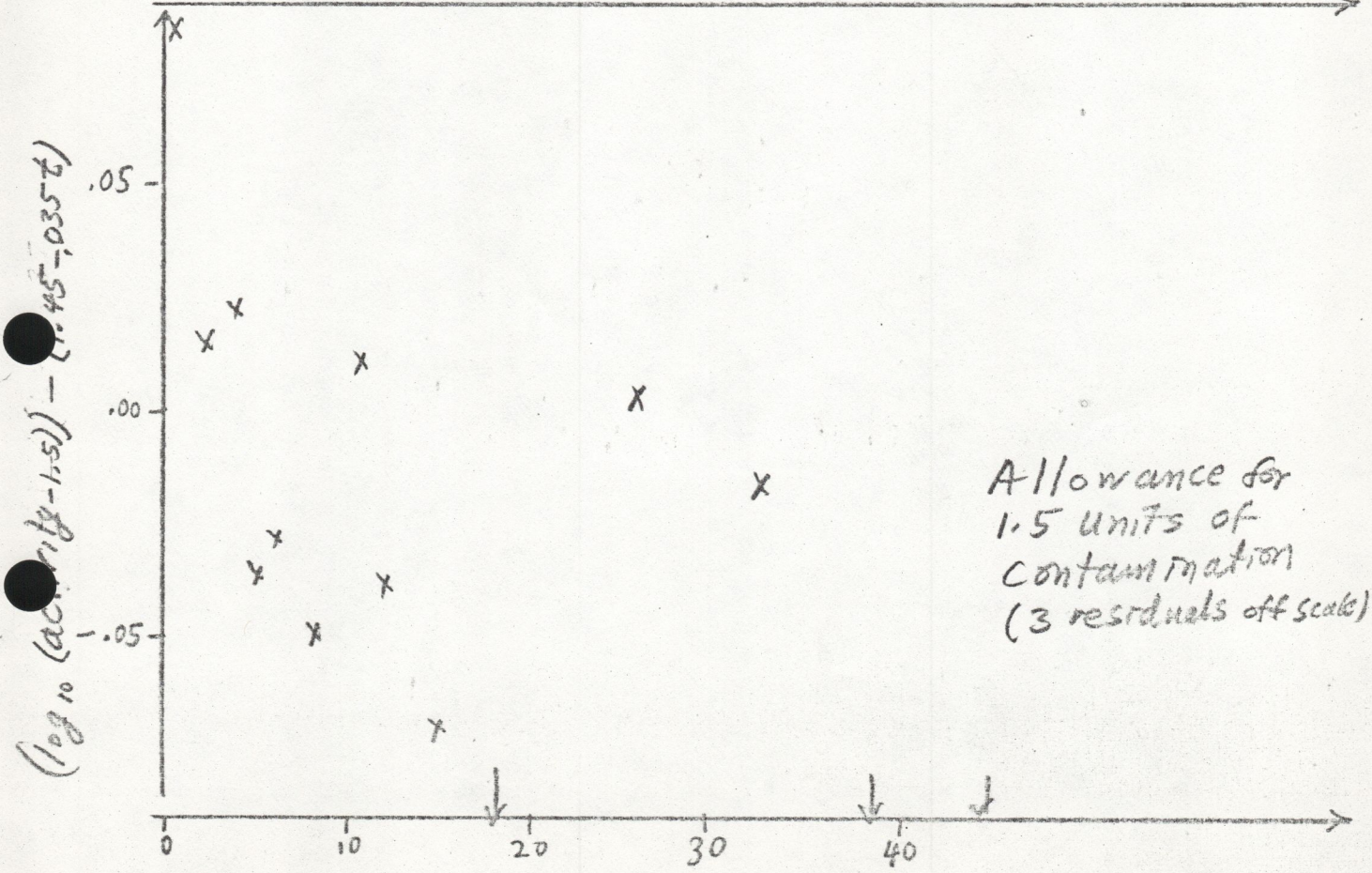
An example of radioactive decay
(logarithmic scale)



plot 34 of chapter 2

Residuals from decay line allowing for long-lived contamination.
(Two versions)





t = Time in days

table 35 of chapter 2

North Carolina's 15 Smallest Counties

(*)	Name	Population (1960) Total	% less than 5 yrs school**	% at least 65	per sq. mile	Families (1959) % less than $3000 income	Local Gov't (1957) % budget for education	Total budget ($1000's)
2874	Tyrrell	4,520	23.3	10.0	11	71.8	54.0	443
2781	Clay	5,526	18.4	11.4	26	70.9	24.1	361
2774	Camden	5,598	10.2	8.7	23	53.1	33.0	206
2765	Hyde	5,765	20.8	13.0	9	68.6	28.8	320
2738	Dare	5,935	16.3	10.7	15	46.3	19.6	372
2668	Graham	6,432	25.7	7.6	22	58.0	41.3	402
2653	Curritack	6,601	19.8	10.1	24	41.8	37.4	310
2536	Allegheny	7,734	16.4	11.6	34	51.3	29.9	328
2446	Swain	8,387	17.6	9.7	16	59.1	54.0	400
2360	Perquimans	9,178	21.0	10.0	35	62.0	53.2	671
2345	Gates	9,254	22.0	9.6	27	60.8	19.1	345
2295	Pamlico	9,850	15.8	9.0	29	52.6	36.2	610
2168	Jones	11,005	18.4	6.2	24	60.3	32.0	472
2128	Polk	11,395	15.2	12.0	49	42.4	24.0	580
2101	Chowan	11,729	19.4	8.1	65	55.5	42.9	659
	(whole state)	((28,784))	(16.5)	(6.9)	(93)	(37.2)	(36.5)	

(*) National (inverse) rank by population.

** Of those 25 or over.

(( )) = Median.

State has 100 counties.

table 36 of chapter 2

South Carolina's 15 Smallest Counties

(*)	Name	Population (1960) Total	% less than 5 yrs school**	% at least 65	per sq. mile	Families (1959) % less than $3000 income	Local Gov't (1957) % budget for education	Total budget ($1000's)
2424	McCormick	8,629	28.2	8.8	23	59.8	62.0	566
2130	Allendale	11,362	34.1	7.9	27	60.1	63.1	1082
2051	Jasper	12,237	37.1	7.5	19	60.2	79.7	1082
2050	Calhoun	12,256	26.1	7.9	33	68.2	76.6	659
1852	Saluda	14,554	17.0	9.7	33	50.6	68.9	656
1763	Edgefield	15,735	23.1	8.0	33	55.3	70.1	785
1720	Bamberg	16,274	26.6	8.3	41	58.5	63.1	1340
1622	Hampton	17,425	31.8	7.2	31	58.0	57.8	1312
1608	Barnwell	17,659	23.2	7.4	32	47.5	82.9	1565
1405	Fairfield	20,713	30.8	7.8	30	54.2	56.8	1596
1356	Abbeville	21,417	23.0	8.7	42	42.1	56.4	1717
1339	Lee	21,832	31.7	6.4	53	68.6	75.8	1671
1211	Dorchester	24,383	23.1	6.6	43	49.6	67.9	1684
1092	Colleton	27,816	29.3	7.5	27	57.9	46.7	3090
1066	Marlboro	28,529	29.6	6.8	59	58.3	63.4	1948
	(whole state)	((34,262))	(20.3)	(6.3)	(79)	(39.5)	( )	

(*) National (inverse) rank by population.

** Of those 25 or over.

(( )) Median.

State has 46 counties.

table 37 of chapter 2

Georgia's 15 Smallest Counties

(*)	Name	Population (1960) Total	% less than 5 yrs school**	% at least 65	per sq. mile	Families (1959) % less than $3000 income	Local Gov't (1957) % budget for education	Total budget ($1000's)
3077	Echols	1,876	28.0	8.5	4	55.6	59.5	122
3051	Quitman	2,432	34.8	9.3	14	70.0	79.5	205
3028	Glascoch	2,672	35.6	56.4	19	61.1	67.9	252
2979	Webster	3,247	34.5	8.7	17	71.2	73.3	225
2978	Schley	3,256	23.0	10.3	20	67.4	69.7	221
2970	Taliaferro	3,370	27.9	13.4	17	68.5	48.8	391
2950	Dawson	3,590	22.4	8.6	17	64.7	71.3	293
2934	Long	3,874	22.7	7.5	10	60.1	52.0	408
2875	Towns	4,538	11.3	10.2	27	63.7	33.4	724
2674	Baker	4,543	33.2	8.9	15	74.1	56.4	622
2873	Clay	4,551	26.8	10.5	20	66.6	59.2	417
2832	Lanier	5,097	30.1	7.5	31	57.4	64.5	686
2810	Charlton	5,313	26.5	6.2	7	44.0	49.1	703
2808	Heard	5,333	22.7	10.7	18	56.4	57.1	652
2805	Wheeler	5,342	29.4	9.5	18	63.7	54.4	418
	(whole state)	((12,038))	(17.6)	(7.4)	(68)	(35.6)	(45.0)	--

(*) National (inverse) rank by population.

** Of those 25 or over.

(( )) Median.

State has 159 counties.

table 38 of chapter 2

Alabama's 15 Smallest Counties

(*)	Name	Population (1960) Total	% less than 5 yrs school**	% at least 65	per sq. mile	Families (1959) % less than $3000 income	Local Gov't (1957) % budget for education	Total budget ($1000's)
2201	Coosa	10,726	18.4	10.7	17	51.5	39.5	1174
2179	Cleburne	10,911	21.5	9.4	19	52.4	48.2	1088
2034	Clay	12,400	14.4	13.0	21	54.0	48.3	1347
1945	Bullock	13,462	32.7	11.7	22	69.4	42.8	1727
1934	Greene	13,600	38.0	9.8	21	74.0	57.9	1514
1876	Lamar	14,271	14.7	11.0	24	51.4	37.5	2507
1870	Bibb	14,357	22.7	9.6	23	54.4	54.7	1405
1828	Winston	14,858	17.3	9.8	24	53.8	55.7	1072
1824	Crenshaw	14,909	23.9	11.2	24	69.5	51.4	1332
1796	Henry	15,286	27.8	9.0	27	63.8	56.0	1313
1794	Washington	15,372	23.3	7.8	14	51.7	54.5	1353
1790	Lowndes	15,417	37.2	9.2	22	72.1	55.4	1243
1729	Fayette	16,148	16.9	10.8	26	54.7	43.0	1425
1715	Cherokee	16,303	16.6	8.9	27	49.1	53.7	1498
1628	Perry	17,358	28.7	10.5	24	69.2	49.4	1689
	(whole state)	((25,738))	(16.3)	(8.0)	(64)	(39.1)	(45.8)	--

(*) National (inverse) rank by population.

** Of those 25 or over.

(( )) Median.

State has 67 counties.

2. The following data was obtained in preparing a standard curve for the determination of formaldehyde by the addition of chromatropic acid and concentrated sulphuric acid and the reading of the consequent purple color on a Beckman Model DU Spectrophotometer at 570 mμ.

Amount of CH_2O Used	Optical Density
0.1	0.086
0.3	0.269
0.5	0.446
0.6	0.538
0.7	0.626
0.9	0.782

Analyze graphically, using at least two graphs. Comment. (Bennett & Franklin, p. 216; from Roberts)

3. The relation between the amount of β-erythrodine dissolved in water and the turbidity of the solution -- as read on a colorimeter -- is not quite as simple. Some data gives:

Concentration (in mg/ml)	Colorimeter reading
40	69
50	175
60	272
70	335
80	390
90	415

Analyze graphically. Comment. (Bennett & Franklin, p. 217, from Wislawski).

4. The amount of carbon in a clay can be measured directly by heating the clay until all the carbon compounds are burned, collecting the carbon dioxide thus formed, and measuring its amount. The amount of carbon can be estimated by combining the amounts of its constituents in a suitable standard way. The results of such measurements on 36 clays from South Devonshire, England are given in table 39. Analyze graphically. Comment. (Bennett and Franklin, p. 218)

5. In a later paper, Meyer and von Schweidler reported the decay of radioactivity for three samples as in table 40. Analyze at least two of these. Comment.

6. The vapor pressure of mercury is stated to be as in table 41. Analyze graphically. Comment.

7. Table 42 gives the % Democratic vote in each of 24 northeastern and central states for the last 12 presidential elections. Plot the following:

(a) 1964 against 1956

(b) 1960 against 1920

(c) 1952 against 1932

(d) any one against another that you think will have a close relationship.

Complete the graphical analysis. Comment.

Problems for Automatic Computation

Not yet prepared.

Theoretical Problems

Not yet prepared.

table 39 of chapter 2

Carbon Content of 36 Clays Measured Directly and Estimated Indirectly

Clay #	Direct Measurement	Indirect Estimate
1	1.53	2.46
2	0.87	1.54
3	0.28	0.70
4	0.27	-0.40
5	3.07	4.82
6	0.25	0.30
7	0.25	0.64
8	0.29	0.78
9	0.12	0.12
10	1.50	2.36
11	1.31	2.14
12	0.31	0.08
13	0.14	-0.01
14	2.98	4.53
15	6.84	9.94
16	2.15	3.68
17	1.35	1.84
18	0.40	0.97
19	4.18	6.14
20	0.22	0.52
21	0.38	0.40
22	0.24	0.46
23	1.79	2.80
24	0.58	2.09
25	6.55	9.68
26	2.54	4.08
27	1.43	2.80
28	2.74	3.93
29	6.08	8.22
30	0.75	0.28
31	0.16	0.35
32	5.06	7.49
33	0.86	1.41
34	0.16	-0.50
35	11.43	15.80
36	0.19	0.18

table 40 of chapter 2

Three Examples of Radioactive Decay

Days*	Activity**	Days*	Activity**	Days*	Activity**
0.8	6.70	0.8	2.82	0.8	2.05
2.8	6.40	2.8	2.34	1.0	2.03
6.9	5.70	4.8	1.90	1.8	1.79
8.9	5.10	6.8	1.80	2.1	1.77
13.1	4.30	11.8	1.34	3.9	1.54
15.2	4.00	13.9	1.24	5.9	1.35
16.8	3.95	16.8	1.03	7.1	1.29
20.1	3.40	19.8	1.00	9.1	1.23
20.8	3.40	23.8	0.80	12.1	1.01
20.9	3.20	31.8	0.55	12.9	0.96
21.9	3.20	32.5	0.11	16.9	0.82
31.1	2.42			19.9	0.68
36.8	2.30			22.8	0.59
43.8	2.11			25.9	0.51
49.8	2.00			33.8	0.405
53.8	1.99			44.1	0.32
58.8	1.98				
65.9	1.90				
73.8	1.80				
87.1	1.65				

* Since start.

** In volts/minute.

From: S. Meyer and E. von Schweidler 1907. Untersuchungen Über radioaktive substanzen. (VIII Mitteilung) Über ein radioaktives Produkt aus dem Aktinium. Sber Ak Wiss Wien Math-Nat Class 116, IIA1, pp. 315-322 (especially 316-317).

table 41 of chapter 2

Vapor Pressure of Mercury

Temperature (°C)	Pressure (mm Hg)
0	0.0004
20	0.0013
40	0.006
60	0.03
80	0.09
100	0.28
120	0.8
140	1.85
160	4.4
180	9.2
200	18.3
220	33.7
240	59.
260	98.
280	156.
300	246.
320	371.
340	548.
360	790.

(Source: Gebhardt, Hertz, Regnault, van der Plaats, and others.)